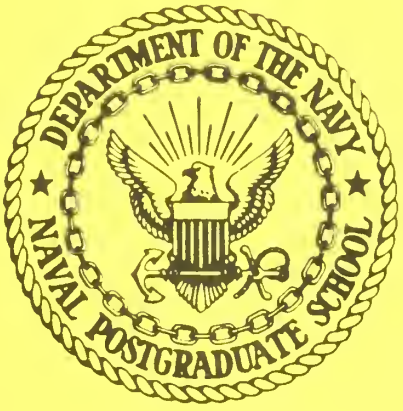
NPS55-83-007

NAVAL POSTGRADUATE SCHOOL

Monterey, California



SIMTBED: A GRAPHICAL TEST BED FOR ANALYZING
AND REPORTING THE RESULTS OF A
STATISTICAL SIMULATION EXPERIMENT

by

P. A. W. Lewis
E. J. Orav
H. W. Drueg
D. G. Linnebur
L. Uribe

April 1983

Prepared for:
Chief of Naval Research
Arlington, VA 22217

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

David A. Schrady
Provost

This work was supported in part by the Office of Naval Research under Grant NR-42-284.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS55-83-007
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE *(and Subtitle)*: SIMTBED: A GRAPHICAL TEST BED FOR ANALYZING AND REPORTING THE RESULTS OF A STATISTICAL SIMULATION EXPERIMENT

5. TYPE OF REPORT & PERIOD COVERED: Technical

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s):
P. A. W. Lewis
E. J. Orav
H. W. Drueg
D. G. Linnebur
L. Uribe

8. CONTRACT OR GRANT NUMBER(s):

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Naval Postgraduate School
Monterey, CA 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
61153N; RR014-05-01
N0001483WR30026

11. CONTROLLING OFFICE NAME AND ADDRESS:
Chief of Naval Research
Arlington, VA 22217

12. REPORT DATE: April 1983

13. NUMBER OF PAGES: 42

14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*:

15. SECURITY CLASS. *(of this report)*: Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT *(of this Report)*:
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*:

18. SUPPLEMENTARY NOTES:

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*:
SIMTBED
Graphics
Simulation
Regression-Adjusted Graphics
Statistical Simulation
Serial Correlation Coefficient
Gamma Distribution

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*:
A graphical test bed in which the results of a simulation experiment can be reported and analyzed is described. The test bed is based on the regression adjusted graphics and estimation methodology developed by Heidelberger and Lewis for regenerative simulation. From the graphics and associated numerics, the experimenter can summarize and see simultaneously relative properties, such as bias, normality and standard deviation, of several estimators of a characteristic of a population for up to 8 sample sizes. The evolution of these properties with sample size is also displayed. The graphics is

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

supported on a line printer to make it and the program portable. The technique is illustrated by two examples, one concerning the effects of changes in data distribution on the behavior of the estimated lag one serial correlation coefficient and the other concerning the relative properties of several estimators of a Gamma distribution.

SIMTBED: A graphical test bed for analyzing and reporting the results of a statistical simulation experiment

P.A.W. Lewis E.J. Orav H.W. Drueg D.G. Linnebur L. Uribe

Naval Postgraduate School
Monterey, California
93940

U.S. Marine Corps
Washington, D.C.

Computer
Applications
Salinas, CA
93908

ABSTRACT

A graphical test bed in which the results of a simulation experiment can be reported and analyzed is described. The test bed is based on the regression adjusted graphics and estimation methodology developed by Heidelberger and Lewis for regenerative simulation. From the graphics and associated numerics, the experimenter can summarize and see simultaneously relative properties, such as bias, normality and standard deviation, of several estimators of a characteristic of a population for up to 8 sample sizes. The evolution of these properties with sample size is also displayed. The graphics is supported on a line printer to make it and the program portable. The technique is illustrated by two examples, one concerning the effects of changes in data distribution on the behavior of the estimated lag one serial correlation coefficient and the other concerning the relative properties of several estimators of a Gamma distribution.

1.0 Introduction

SIMTBED is a graphical display program that can be used via a simulation on a digital computer to (i) explore the distribution of a statistical estimator for a given sample size, (ii) to compare the properties of that distribution when the estimator is calculated for various sample sizes, and (iii) to contrast those properties under different estimation conditions. Those conditions are controlled by the experimenter but, most commonly, they will entail competing estimation procedures (e.g. maximum likelihood versus methods of moments, or jackknifed versus not jackknifed). The program is flexible enough to accommodate the imagination of most users and, in one of the examples, we also consider the effects of changes in the underlying distribution of the data.

One salient feature of the program is that it uses the same batch of simulated random variables (e.g. Normals) to explore the properties of all the estimators at various sample sizes. This is done for economy of computer time and could be important on slow computers; the price paid is that the analytical analysis provided by SIMTBED of its graphical output is performed on correlated samples.

To use the program it is necessary only to define the optional input parameters, supply the simulated random variables, and provide the Fortran functions which, when passed the data and subsample size, transform (if desired) the data subsample and compute the desired statistics. SIMTBED itself will subdivide and feed the data properly into the functions, produce boxplots and summary statistics, and compute regressions for the mean and variance of each estimator based on inverse subsample size. Up to three estimators can be used with the option to

produce equally scaled graphs for all the statistics.

The features of the program are more easily demonstrated by example rather than explanation and so we will proceed directly to two applications. The first application refers back to a simulation study done by Cox (1966) looking at the behavior of the estimated first order serial correlation coefficient, Fisher's z-transform of the estimated correlation, and the 2-fold jackknifed estimate of the correlation for i.i.d. Normal(0,1), $\chi^2(1)$ and Lognormal(0,1) data. The jackknife was originally proposed by Quenouille (1948) for the purpose of removing bias from the correlation estimate. The second application considers the problem of estimating the shape parameters for a highly skewed Gamma(.25) and a nearly Normal Gamma(5.0) sample using m.l.e., method of moments, 4-fold jackknifed m.l.e., and 4-fold jackknifed method of moments as the competing estimators.

Technical details concerning the SIMTBED software, not essential to interpreting and appreciating the output, can be found in Linnebur (1982), and an application to the analysis of output in a regenerative simulation can be found in Heidelberger and Lewis (1981).

2.0 Calculation of the First Serial Correlation Coefficient

It is known that for an independent sample from a population with finite variance, the distribution of the serial correlation coefficient (Anderson and Walker, 1964) is asymptotically Normal with mean zero and variances 1/n, where n is the sample size. If the population is i.i.d Normal then the bias is exactly -1/n. Since those asymptotic properties are frequently used as approximations in tests of significance, it is important to know how valid the approximation would be in small samples

from a variety of distributions. We will look at that question in the next two sections and then go on to consider two alternative measures of correlation, Fisher's z-transform and the 2-fold jackknifed estimate of the correlation. Their ability to reduce bias and/or induce Normality will be examined against other changes in the distribution of the estimators, particularly variance inflation. A simulation study, without graphics, of some of these problems was conducted by Cox (1966). He did not consider the jackknifed estimate.

2.1 SIMTBED Output for Serial Correlation

Figure 1(a) shows the simulated distribution and sample properties of the serial correlation coefficient estimate

$$r_n = \frac{n \sum_{j=1}^{n-1} (X_j - \bar{X}_1)(X_{j+1} - \bar{X}_n)}{(n-1) \sum_{j=1}^{n} (X_j - \bar{X}_0)^2},$$

where $\bar{X}_0 = \sum_{j=1}^{n} X_j/n$, $\bar{X}_1 = \sum_{j=1}^{n-1} X_j/(n-1)$, and $\bar{X}_n = \sum_{j=2}^{n} X_j/(n-1)$

for various sub-sample sizes $n=n_i$. This definition matches that used by Anderson and Walker (1964). We consider first subsamples of size $n_1 = 10$, and then of size $n_2 = 20$, $n_3 = 30$, $n_4 = 40$, $n_5 = 50$, $n_6 = 75$, $n_7 = 100$ and $n_8 = 150$, successively. For each subsample size the input sample of N = 5000 simulated Normal(0,1) random variables is divided into as many full subsamples of size n_i as possible, and the serial correlation is computed for each of the $\lfloor N/n_i \rfloor$ subsamples of size

n_i . The entire procedure is then replicated $M = 10$ times, each time with a new simulated sample of $N = 5000$ Normal(0,1) variables.

After all M replications have been run, all the estimates of serial correlation for each subsample size are grouped together and their simulated distribution is presented via a boxplot and summary statistics (see e.g. Fig. 1(a)). The boxplot follows the standards discussed in Mosteller and Tukey (1977) with the median denoted by a + within the box, the mean by a * within the box, the outliers by 0's, and the far outliers by *'s beyond the whiskers. The summary statistics include the sample mean, sample standard deviation, estimated standard deviation of the sample mean (i.e. sample standard deviation/sqrt$\left(M\left\lfloor N/n_i \right\rfloor\right)$, sample skewness and sample kurtosis of the correlation estimates.

Looking at the output, the first (leftmost) boxplot in the graph in Figure 1(a) shows the distribution of

$$(\#\ \text{Replications}) \times \left\lfloor \frac{(\text{Simulation Sample Size})}{(\text{Subsample Size})} \right\rfloor = 10 \times \left\lfloor \frac{5000}{10} \right\rfloor = 10 \times 500 = 5000$$

estimates of serial correlation from independent subsamples of size $n_1 = 10$. Summary statistics for the boxplot can be found below the graph in the column labeled "Subsample Size 10", so that the average serial correlation is -.1074, and the estimated standard deviation is .2996. The estimated standard deviation of the serial correlation estimate is $.2996/\sqrt{(5000)} = .00424$. Recall that this refers to correlation estimates based on subsamples of size 10.

Since the X-axis of the graph represents subsample size, the last (rightmost) boxplot shows the distribution of

$$10 \times \left\lfloor \frac{5000}{150} \right\rfloor = 10 \times 33 = 330$$

estimates of serial correlation from independent subsamples of size $n_8 = 150$. Although the 330 estimates are independent of each other, they are not independent of the 5000 estimates that comprise the first boxplot since the same data (divided and processed in different ways) was used for both. Summary statistics show that the average correlation has dropped to -.007372, indicating the fall off in bias, and the standard deviation has dropped to .07822, indicating the greater precision with which the correlation can be estimated when 150 points, rather than 10, are available.

In order to quantify the changes that are occurring in the mean and variance of the distribution of the estimator as subsample size changes, SIMTBED performs two types of regressions. The first regression is on the averages and is done after each replication, using the average serial correlation for that replication, $\bar{r}_{n_i}$, as the dependent variable. Inverse powers of the subsample size serve as the independent variables. For Figure 1(a) the degree of the regression was chosen to be D=3 so, for each replication, the equations we attempt to fit by least squares are:

$$\bar{r}_{n_i} = a_0 + \frac{a_1}{n_i} + \frac{a_2}{n_i^2} + \frac{a_3}{n_i^3} \qquad \text{for } i = 1,2....8.$$

This form anticipates the general asymptotic expansion

$$E\,(\hat{\theta}(n)) = \theta + \frac{\alpha_1}{n} + \frac{\alpha_2}{n^2} + \frac{\alpha_3}{n^3} + ...$$

which holds true in the current situation with $\theta = 0$ and (in the Normal case) $\alpha = -1$ (see Cramer, 1948, for general results of this type).

Values of a_0, a_1, a_2, and a_3 are calculated after each replication, averaged across the M replications to get $\bar{a}_0$, $\bar{a}_1$, $\bar{a}_2$, and $\bar{a}_3$, and then the averages are reported below the summary statistics on the line "Mean of Regression on Averages - Coefficients". We find that

$\bar{a}_0 = -.000272$ and $\bar{a}_1 = -1.03074$, both close to their respective theoretical counterparts of zero and -1.

Because we have 10 replications and therefore 10 independent values of each of a_0, a_1, a_2, and a_3, we can also estimate the variances and standard deviations of $\bar{a}_0$, $\bar{a}_1$, $\bar{a}_2$, and $\bar{a}_3$ across replications. These values are presented on the two lines immediately below the coefficients. For instance, the estimated s.d. of the estimate $\bar{a}_0 = -.000272$ of a_0 is .003892.

The regression line for the mean value of the estimator is presented visually in the graph as a dotted curve. The estimated asymptote (i.e. $\bar{a}_0$) is printed with a dashed line wherever it does not coincide with the regression line. Bias, therefore, can be viewed as the difference between those two lines.

The second regression referred to above is done after all replications have been run and the variances of the estimators at each subsample size have been calculated. (Note that the standard deviations, not the variances, are presented in the summary statistics.) It should be recalled from previous discussion that these variances, as well as all measures in the summary statistics, are based on the grouping together of the serial correlations <u>from all replications</u>, at each subsample size. This is in contrast to the the procedure for the regression on the means, where average correlations are computed for each subsample size <u>for each replication</u>. In the case of the variances, we have 8 equations:

$$\hat{\mathrm{Var}}(r_{n_i}) = \frac{b_0}{n_i} + \frac{b_1}{n_i^{3/2}} + \frac{b_2}{n_i^2} + \frac{b_3}{n_i^{5/2}}, \qquad i = 1,2.....8,$$

which we fit by least squares in order to estimate the coefficients β_0, β_1, β_2, and β_3 in the presumed asymptotic expansion

$$\text{Var}\ (\hat{0}\ (n)) = \frac{\beta_0}{n} + \frac{\beta_1}{n^{3/2}} + \frac{\beta_2}{n^2} + \frac{\beta_3}{n^{5/2}} + \ldots.$$

This expansion holds for the variance of the estimated serial correlation coefficient for independent data. Usually it will be β_0 in which we are most interested since β_0 is used in computing asymptotic relative efficiencies of estimators. For independent data with finite variance, we know that $\beta_0 = 1$. The computed values of b_0, b_1, b_2, and b_3, are presented on the line labeled 'Regression on Variance - Coefficients'. Notice that $b_0 = .7438$ is close to the theoretical value of 1.

The final two numbers on Figure 1(a), YMIN and YMAX, simply show the scale of the vertical axis. Because the SIMTBED program option to put Figures 1(a), 1(b) and 1(c) on the same scale was in effect, it may be that no boxplot in a given Figure (eg. Figure 1(b)) requires the full range of Y-values.

In order to produce Figure 1(b), the Normal(0,1) data that went into Figure 1(a) was squared to create longer tailed $\chi^2(1)$ random variables. The output is entirely analogous to that for Figure 1(a). Similarly, for Figure 1(c), the Normal(0,1) data was exponentiated in order to create Lognormal(0,1) data and to produce analogous graphical output. The indication is that the distribution of the sample serial correlation is robust with respect to the population distribution.

The features of the SIMTBED output will become clearer when they are associated with the various properties of the correlation estimator. First, however, a few technical comments concerning the regressions are necessary.

2.2 Some Comments on the Regressions

Two types of problems, numerical and statistical, can occur when

attempting to fit the two sets of regression equations presented in Section 2.1.

First, there is the question of numerical stability when the independent variables, $\{1, n_i^{-1}, n_i^{-2}, n_i^{-3}\}$ or $\{n_i^{-1}, n_i^{-3/2}, n_i^{-2}, n_i^{-5/2}\}$ decrease geometrically. If we attempt to form X^TX, where X is the respective design matrix and X^T is the transpose of X, we get values that range from 8 (assuming 8 subsample sizes) to $\sum_{i=1}^{8} n_i^{-6}$ for the regression on the means, and $\sum_{i=1}^{8} n_i^{-2}$ to $\sum_{i=1}^{8} n_i^{-5}$ for the regression on the variances. Experience has shown that attempts to solve systems with such extremes in the X^TX matrix produce erroneous results. Consequently, SIMTBED scales the design matrices by multiplying each entry of X by $\underset{i}{Max}(n_i)$ raised to the proper power so that no entry becomes too small. The standard Choleski decomponsition (see Dahlquist, Bjorack and Anderson, 1974) is then used to fit the equations, and the coefficients are properly rescaled before they are reported. This procedure produces numerically reliable results.

The second problem concerns the breakdown of statistical assumptions in our regression models. It has already been pointed out in Secion 2.1 that the two sets of dependent variables:

(1) the $\bar{\theta}(n_i)$ when considering the regression on the means;

(2) the $s^2(n_i) = \sum_{j=1}^{M\lfloor N/n_i \rfloor} (\hat{\theta}_j(n_i) - \bar{\bar{\theta}}(n_i))^2 \Big/ M\lfloor N/n_i \rfloor$,

where $\bar{\bar{\theta}}(n_i)$ is the mean across the M replications, when considering the regression on the variances,

Table 1

Entries in the table are the estimated correlations between the estimated variances of the r_{n_i} at different subsample sizes: Corr $(s^2(r_{n_i}), s^2(r_{n_j}))$ for i=1,...,8, j=1,...,8.

i / j	1	2	3	4	5	6	7	8
1	1.00	.49	.46	-.26	.18	-.17	.14	.01
2	.49	1.00	.40	.55	.11	.38	.38	-.03
3	.46	.40	1.00	.23	.23	.44	.21	.29
4	-.26	.55	.23	1.00	.42	.86	.57	.35
5	.18	.11	.23	.42	1.00	.71	.43	.59
6	-.17	.38	.44	.86	.71	1.00	.45	.53
7	.14	.38	.21	.57	.43	.45	1.00	.72
8	.01	-.03	.29	.35	.59	.53	.72	1.00

Recall that r_n is the estimated serial correlation for a simulated Normal(0,1) subsample of size n. Also, the estimated correlations shown above were computed using 10 values (replications) of $s^2(r_{n_i})$ and $s^2(r_{n_j})$ for each i and j.

Table 2

A comparison of the estimated variance of $s^2(r_{n_i})$ with the approximate theoretical variance of $s^2(r_{n_i})$ and with the approximately equivariant scaled versions, $n_i^{-.5}\ s^2(r_{n_i})$. All entries have been multiplied by 10^5.

n_i =	10	20	30	40	50	75	100	150
$\hat{\text{Var}}\,(s^2(r_{n_i}))$	.177	.150	.204	.079	.047	.031	.049	.022
Approx. Theoretical Var $(s^2(r_{n_i}))$	.400	.200	.133	.100	.080	.053	.040	.027
$\hat{\text{Var}}(n_i^{-.5}\ s^2(r_{n_i}))$	1.77	2.99	6.12	3.18	2.33	2.33	4.88	3.35

The estimated variances of $s^2(r_{n_i})$ and $\sqrt{n_i}s^2(r_{n_i})$ were calculated using 10 independent replications of $s^2(r_{n_i})$.

are not independent over i since all are based on the same simulated data. The extent of the dependence is demonstrated by the correlation matrix in Table 1. Entries in that table show the estimated correlation between $s^2(n_i)$ and $s^2(n_j)$ for all i and j, where the estimation was done by repeating the SIMTBED experiment with 10 different batches of 50,000 simulated random variables. Since only 10 values went into each correlation calculation, the table is only accurate to within approximately $\pm 2/\sqrt{10} = .632$. We see some indication of positive correlation, especially when i and j are close, but the lack of independence is not severe enough to hurt the regression results for either the estimated means or variances significantly.

A second assumption, implicit in any regression, is that the dependent variables have equal variances. This condition holds true for the means, which can be shown to satisfy

$$\text{Var}\ (\bar{\theta}(n_i)) = \frac{M}{N}$$

independently of i. The estimated variances, however, are not equivariant and, if we assume the $\hat{\theta}_j(n_i)$ to be approximately Normally distributed so that

$$M\left\lfloor N/n_i \right\rfloor \sum_{j=1} (\hat{\theta}_j(n_i) - \bar{\bar{\theta}}(n_i))^2$$

is approximately proportional to a Chi-squared random variable, with $M\left\lfloor N/n_i \right\rfloor - 1$ degrees of freedom, we can compute

$$\text{Var}\ (s^2(n_i)) \cong \frac{2}{MNn_i - n_i^2} .$$

To correct this problem of unequal variances SIMTBED scales the $s^2(n_i)$ by $\sqrt{n_i}$ so that

$$\text{Var}\left(\sqrt{n_i}\, s^2(n_i)\right) \cong \frac{2}{MN - n_i} \cong \frac{2}{MN}$$

since $MN >> n_i$. The design matrix is scaled accordingly and the values b_0, b_1, b_2, and b_3 discussed in Section 2.1 are reported.

Table 2 shows the effects of the rescaling by presenting first the estimated variances of the $s^2(n_i)$, where the estimation is done by repeating SIMTBED for 10 batches of 50,000 simulated data points. These estimated variances decrease as n_i increases, closely paralleling the second line of Table 2 which has the approximate theoretical values (i.e. $2/(MNn_i - n_i^2)$). The final line of Table 2 shows the estimated variances of the rescaled $s^2(n_i)$, i.e. the $\sqrt{n_i}\, s^2(n_i)$, which, as expected and hoped, show a more constant variance with i.

Although future versions of SIMTBED will include more sophisticated regression routines and the ability to generate independent samples at each subsample size, the current version is quick, usable, and accurate for most situations.

2.3 Interpreting the Serial Correlation Results

Returning to Figure 1(a) which shows the simulated distribution of the serial correlation coefficient from independent, Normal(0,1) data, the following comments summarize the most striking features:

(a) The boxplots appear very symmetric at all subsample sizes with nearly equal numbers of outliers at either tail and with mean and median coincidental. This observation is confirmed by the estimates of skewness in the summary statistics. Kurtosis is mildly negative at small subsample sizes but, overall, asymptotic Normality seems to take hold rather quickly. Note however, that at $n_i = 10$ there are only 3 outliers in a

sample of size 5000. This is consistent with the estimated kurtosis of -0.424, showing that the distribution is quite nonnormal.

(b) The average serial correlation is negative for small subsamples. This is demonstrated by the dotted regression curve which starts at approximately -.10 and levels off near 0 for subsamples greater than about 85. The dashed asymptote of -.000272 is very close to the theoretical value of 0, and the mean values in the summary table closely reflect the bias of - 1/n.

(c) The standard deviations of the simulated distributions are very close to the asymptotic values of $n_i^{-0.5}$, although the lead coefficient in the regression on the variances, $b_0 = .743756$, is not as close to the theoretical value of 1 as we would hope. When SIMTBED is repeated 10 times with 10 different batches of simulated data, we find an average value for b_0 to be 1.0604, with a standard deviation for b_0 of .307. The estimation procedure for b_0, therefore, remains valid, but the estimate itself is highly variable.

The agreement between the simulated and the theoretical, asymptotic values of the bias and variances was discovered previously by Cox (1966). SIMTBED has now allowed us to automatically look at a broader range of subsample sizes and to see, through boxplots and estimates of skewness and kurtosis, a fuller picture of any changes in the distribution of the estimator. We can be satisfied that estimates of serial correlation do behave approximately as Normal(-1/n, 1/n) random variables when the underlying data is Normal(0,1).

If the lead terms in the expansions of the mean and variance of the estimated correlation coefficient (ie. a_0, a_1, and b_0) had been unknown, we would also have a fairly good idea now of what they were.

When the underlying data is χ^2_1, Figure 1(b) confirms Cox's observation that the bias is relatively uneffected but, for small subsamples, the standard deviation is smaller than the expected $n^{-0.5}$. Unlike Figure 1(a), there is a pronounced skewness in the boxplots in Figure 1(b) with many more outliers at the positive end, and with the mean higher than the median at the first four subsample sizes. The problem of suppressed variance seems cured at $n_7 = 100$ and $n_8 = 150$, but the skewness remains and could cause problems in tests of significance.

Figure 1(c), which is based on an underlying batch of simulated Lognormal(0,1) data, shows a slight exaggeration of the effects in Figure 1(b). The standard deviation is more suppressed and does not attain the theoretical level by $n_8 = 150$. The positive skewness is more pronounced and kurtosis does not approach the theoretical value of 0.

Overall, the effects of long-tailed data on the distribution of the serial correlation coefficient can be summarized as follows:

(i) Bias is not significantly affected and remains at approximately $-1/n$.

(ii) The variance of the distribution of the serial correlation coefficient is reduced by longer-tailed data.

(iii) Positive skewness is created in the distribution.

(iv) Kurtosis may become positive at large subsample sizes.

(v) For long-tailed data (i.e. Lognormal), a subsample size of 150 is not large enough to insure asymptotic Normality.

2.4 SIMTBED Output for the z-Transform of the Correlation

Fisher's z-transform of the estimated correlation coefficient is defined by:

$$Z_n = \frac{1}{2} \log \frac{1 + r_n}{1 - r_n} ,$$

where r_n is the estimated serial correlation presented in Section 2.1. The transformation is intended to make the distribution of the Z_n more Normal than that of the r_n. When the same SIMTBED experiment described in Section 2.1 is run using Z_n as the estimator instead of r_n, we get the results shown in Figures 2(a), 2(b) and 2(c). It should be noted that the scale of the boxplots here has been forced to be approximately comparable to the scale for the boxplots in Section 2.1. This is done by suppressing outliers that are more than 1.5 interquartile distances beyond the quartiles of the boxplot. If we had allowed the data to scale the boxplots, we would have seen a much wider range on the vertical axis because the Z_n are not restricted to the limits of -1 to $+1$ and because there is one far outlier at -3.8. In this type of "reduced graphics", we still see the number of outliers that fall beyond the allowable range through the numbers at the ends of the boxplots, but we do not see their actual locations.

Figure 2(a) shows the distribution of the z-transformed correlation coefficients when we use simulated Normal(0,1) data. At each subsample size, the mean and standard deviation are close to the theoretical n^{-1} and $n^{-1/2}$ respectively. The skewness and kurtosis at subsample size $n_1 = 10$ are far from the theoretical Normal distribution values of 0 and 0, reflecting partly the one far outlier at -3.8 and partly the negative skew in the remainder of the Z_{n_1}'s. For other subsample sizes, there is no strong evidence to contradict the assumption of approximate Normality.

The relationship between Figure 2(b) and 2(a) is similar to that between 1(b) and 1(a). Figure 2(b), which is based on simulated χ^2_1

data, shows (a) bias that is the same as for the transformed correlations based on Normal data, (b) slightly suppressed variances, particularly at small subsample sizes and (c) positive skewness which persists at large subsample sizes. In addition, there are signs of positive kurtosis at small subsample sizes.

Figure 2(c) is based on Lognormal(0,1) data and shows high values of skewness and kurtosis at almost all subsample sizes. Approximate Normality seems an unwarranted assumption. In fact, the kurtosis is converging very slowly to its asymptotic value of 0.

In general, using the z-transform does not help with Normality assumptions, especially when dealing with long-tailed distributions.

2.5 SIMTBED Output for the 2-Fold Jackknife of the Correlation

The final Figures, 3(a), 3(b) and 3(c), deal with the 2-fold jackknife estimate of correlation. Again, the figures are reduced graphics with scaling comparable to that of the boxplots of Sections 2.3 and 2.4. To define the estimator, we start with a given subsample of size n, compute the serial correlation for the first $\lfloor n/2 \rfloor$ points and call it $r_1(n/2)$, compute the serial correlation for the second $\lfloor n/2 \rfloor$ points and call it $r_2(n/2)$ and compute the serial correlation for the entire subsample of n points and call it $r_0(n)$. Each computation follows the formula in Section 2.1. The three estimators are then combined to form two pseudo-values,

$$r_1^*(n) = 2r_0(n) - r_1(n/2)$$

and

$$r_2^*(n) = 2r_0(n) - r_2(n/2) ,$$

and the final jackknife estimator for that subsample is defined as

$$\tilde{r}(n) = \frac{r_1^*(n) + r_2^*(n)}{2} .$$

Although a jackknife estimator may have many favorable properties, we are concerned here primarily with its ability to remove bias, hopefully without inflating the variances of the estimator and/or inducing nonnormality.

Figure 3(a), based on simulated Normal(0,1) data, shows nearly complete removal of bias, even at small subsample sizes. The cost of the bias reduction is reflected in an increase of nearly 50% in the standard deviation of the correlation estimate for subsample size 10, and lesser relative increases at larger subsample sizes. There is also an indication of a positive skew for small subsample sizes, and the problem that the jackknife estimator need not fall into the -1 to +1 range which is desirable for a correlation coefficient estimate.

When using simulated χ_1^2 data as in Figure 3(b) , or simulated Lognormal(0,1) data as in Figure 3(c), there is again no problem with bias. Variance inflation, though it exists at small subsample sizes, is not as large as when Normal(0,1) data is used. The distributions of the jackknifed correlations show very pronounced positive skews, however, as well as positive kurtosis. These two problems are worse for the longer-tailed Lognormal data.

Overall, the jackknife estimator is very successful at removing bias but the costs include variance inflation, which can be severe at small subsample sizes, plus increased positive skewness and kurtosis when the estimates are based on data from longer-tailed distributions.

2.6 Comparison of the Three Estimates of Correlation

For Normal(0,1) data, the distribution of the usual correlation coefficient displayed in Figure 1(a) behaves very much as theoretical

asymptotic calculations would predict, even at small subsample sizes. This makes it possible to correct for bias in the estimator and to perform tests of significance. Use of Fisher's z-transform, as illustrated in Figure 2(a) does not seem necessary since it does not significantly improve the approximate Normality of the estimator. The jackknife estimator in Figure 3(a) may be valuable if a direct, unbiased estimator is needed but the inflated variance of the jackknife estimator may limit the usefulness of the estimate as well as make any tests of significance too conservative.

When the underlying data comes from a longer-tailed distribution, the usual correlation coefficient in Figures 1(b) and 1(c) retains a predictable bias term, although the variance of its distribution is slightly depressed and the skewness and kurtosis becomes positive, even for subsamples as large as 150. This means that it is still possible to estimate the correlation accurately, but tests of significance fall on shakey assumptions of Normality. The z-transform in Figures 2(b) and 2(c) does little to firm up those assumptions and, in some cases, makes the situation worse. As in the case of Normal data, the 2-fold jackknifed correlation in Figures 3(b) and 3(c) is bias-free but follows a fairly nonnormal distribution which would invalidate significance testing.

All of the preceding observations and conclusions flow immediately from the nine Figures presented so far. Further studies could easily be done through SIMTBED, looking at larger subsample sizes, correlated data, i.e. $\rho \neq 0$ and alternative marginal distributions. For demonstration purposes, though, it is better to proceed to our second application.

3.0 Estimating the Shape Parameter for a Gamma Distribution

As a second application of SIMTBED, we will consider a problem

which has received much less statistical attention; asymptotic results are summarized in Cox and Lewis (1966, Ch.3) and Johnson and Kotz (1970, Ch.17). We want to estimate the shape parameter, K, for a Gamma distribution, where the Gamma density is given by

$$f(x) = \left(\frac{K}{\mu}\right)^{K} \frac{x^{K-1}}{\Gamma(x)} e^{-Kx/\mu} \qquad x \geq 0;\ K > 0;\ \mu > 0$$

$$= 0 \qquad x < 0 .$$

Notice that the mean of this distribution is μ, not K/μ as in some differently parameterized versions of the Gamma density. For the data that will be simulated for use in SIMTBED we will use $K = 5$ and $\mu = 1$ and $K = 0.25$ and $\mu = 1$. The closer the mean of our estimate is to 5 or 0.25, the better (in terms of bias) is our estimation procedure. Other factors such as the variance and Normality of the estimator will of course also have influence in the determination of a prefered estimator; the bias and variance could be combined into m.s.e.

Section 3.1 will compare the commonly used maximum likelihood estimator which is mildly difficult to compute to the competing method of moments estimator which is very simple to compute. Both procedures result in asymptotically Normal estimators (Cramer, 1948) but the m.l.e. is usually prefered because of its favorable asymptotic relative efficiency (Cox and Lewis 1966). Through SIMTBED, though, we will see that for small subsamples the estimated variances of the two estimators of K are not as far apart as asymptotic results lead us to believe. In addition, the bias that appears in both estimators is smaller for the moment estimator.

In Section 3.2 we will use a four-fold jackknife of both the m.l.e. and moments estimators to successfully remove the bias. What is remarkable is that unlike the jackknifing of the serial correlation, there is

little or no cost in terms of variance inflation and nonnormality for the jackknifed moment estimator. When $K = .25$, we will see in Section 3.3 that the jackknifed m.l.e. dominates the other three estimators at all subsample sizes when considering the mean, variance, and Normality of the estimator.

3.1 Maximum Likelihood and Moment Estimators of K

Figure 4(a) is very similar in format to the figures that have already been presented for the correlation example except that:

(1) The estimator whose distribution is being displayed is the maximum likelihood estimator of K, the shape parameter of a Gamma(5) population. We denote the estimator, computed from a simulated subsample of size n, by $\hat{K}(n)$ and define it to be the solution of the equation:

$$n\,[\,\log \hat{K}(n) - \Psi(\hat{K}(n))] = n \log \sum_{i=1}^{n} X_i/n - \sum_{i=1}^{n} \log X_i \,,$$

where the X_i are the simulated Gamma(5) random variables and $\Psi(.)$ is the digamma function (Cox and Lewis, 1966).

(2) The eight subsample sizes which we will be looking at are $n_1 = 33$, $n_2 = 50$, $n_3 = 71$, $n_4 = 100$, $n_5 = 125$, $n_6 = 166$, $n_7 = 250$ and $n_8 = 500$. Note the difference between the n_i's in the previous example and these n_i's. Since these are larger we will not see much small sample detail, but we will see some of the asymptotic ($n = 500$) effects coming in.

(3) At each subsample size we will work with $M^* = 20$ independent replications of $N^* = 2500$ simulated Gamma(5) random variables, instead of the $M = 10$ replications of $N = 5000$ variables used previously.

The total number of independent simulated random variables across replications remains constant at the program maximum of 50,000. Hence, the boxplot at subsample size 50 in Figure 4(a) represents the distribution of $M^* \lfloor N^*/50 \rfloor = 1000$ estimates of $\hat{K}(50)$ just as the boxplot at subsample size 50 in Figure 1(a) represents the distribution of $M \lfloor N/50 \rfloor = 1000$ estimates of $r(50)$. As long as the product, M x N, remains constant, the only effect of changing the number of replications is, up to rounding in N/n_i, to change the results in the regression on the averages. By using $M^* = 20$ and $N^* = 2500$, SIMTBED reports regression coefficients averaged over 20 replications, but, within each replication, the dependent variables are averages over just $\lfloor 2500/n_i \rfloor$ values of the estimator.

(4) The boxplots are presented using the reduced graphics option. In this option any extreme outliers (i.e. those beyond 1.5 interquartile distances) are included as a count at the tail of each boxplot. This option was chosen in order to give more graphical weight to the body of the distributions and the fall-off in the bias. Limited printer resolution makes it impossible to show details in the body and the tails of the distributions if there are many straggling outliers. In the case of very extreme outliers, no detail would be seen in the body of the boxplot without the reduced graphics option.

Figure 4(b) looks at the distribution of the moment estimator of K, the shape parameter of a Gamma(K) population:

$$\tilde{K}(n) = (n-1)\ \bar{X}^2 \ /\ \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad ,$$

where $\bar{X} = \sum_{i=1}^{n} X_i/n$, n is the subsample size, and the X_i are the simulated Gamma(5) random variables. The SIMTBED options and para-

meters mentioned in (2), (3) and (4) preceding are also in effect here.

The two figures, 4(a) and 4(b), show a very pronounced bias in both estimation procedures, although the moment estimator is slightly closer to the unbiased value of 5. As expected, the standard deviation of the m.l.e. is lower than that of the moment estimator although the relative difference at small subsample sizes, for instance 1.448 versus 1.482 at $n_1 = 33$, may not outweigh the increase in bias with the m.l.e. At larger subsample sizes, the relative difference is close to the theoretical asymptotic relative efficiency of .78 (i.e. .91 at $n_7 = 250$)

Both estimators also show distributions with positive skewness and kurtosis that decrease to the asymptotic 0 levels as subsample size increases. The asymptotics appear to take hold more quickly for the moment estimator than for the m.l.e.

In summary, SIMTBED shows that the m.l.e. is indeed better than the moment estimator in terms of variance, but not as good for small sample sizes as asymptotic results would lead us to believe. In the other areas of bias and asymptotic Normality, the moment estimator would have to be preferred.

3.2 4-Fold Jackknifed Estimators of K

Figures 4(c) and 4(d) show the distributions of the 4-fold jackknifed m.l.e. of K and 4-fold jackknifed moment estimator of K, respectively. A 4-fold jackknife estimator is similar to the 2-fold jackknife estimator described in Section 2.5 except that there are 4 pseudo-values that come out of dividing each subsample into fourths. More details can be found in Mosteller and Tukey (1977).

The purpose of the jackknife is to remove the conspicuous bias

observed in Figures 4(a) and 4(b). This goal is seen to be accomplished in Figure 4(c) and 4(d) and we can also note smaller values of skewness and kurtosis, indicating a quicker approach to asymptotic Normality. The skewness and kurtosis of the jackknifed moment estimator are the lowest, at small subsample sizes, among all estimators. The variance of the jackknifed moment estimator is also only slightly inflated, as is the variance of the jackknifed m.l.e..

All told, the jackknifed moment estimator, because of its lack of bias, small variance, and low skewness and kurtosis, would be the method of choice if estimation of K or significance testing was the goal.

3.3 Results for K = 0.25

In Figures 5(a), 5(b), 5(c) and 5(d) we show similar results to those discussed above for the case K = 5.0, but using K = 0.25. The fact (Cox and Lewis, 1966, Ch.3) that the m.l.e. estimate is much more efficient than the moment estimate is graphically illustrated. What is new is the effect of jackknifing: bias is reduced without the sacrifice of variance inflation or nonnormality.

Further comparisons and interpretations are similar to those done for the case K = 5.0, and are left to the reader.

4.0 Conclusions

Simply by providing SIMTBED with the desired estimators, we have been able (a) to explore in depth the effects of changes in data distribution and of different estimation procedures on the calculation of the serial correlation coefficient, and (b) to compare four different ways to estimate the shape parameter in a highly skewed Gamma population.

The graphics and numerical output combine to let us see and quantify distributional changes that occur as subsample size grows. We can see bias fall away, variance shrink, and skewness disappear as the estimator approaches asymptotic Normality. Terms in the asymptotic expansion of the mean and variance of the estimator are automatically calculated and can be used to compare different estimators.

Ease of use and portability, however, remain as SIMTBED's most important features, and will hopefully inspire users to try more diverse and extensive simulation experiments.

Other graphical displays besides running boxplots can be used; some alternatives are given in Lewis (1972), Heidelberger and Lewis (1981) and Devlin, Gnanadesikan and Kettenring (1981).

5.0 Availability

SIMTBED is at present only available in a version run on the IBM 3033. However, since the program uses standard FORTRAN and is independent of any software packages, conversion should be simple. Versions for VAX machines will be tested shortly.

References

Anderson, T.W. and Walker, A.M., (1964). "On the Asymptotic Distribution of the Autocorrelation of a Sample from a Linear Stochastic Process", Ann. Math. Statist., 35, pp. 1296-1303.

Cox, D.R., (1966). "The Null Distribution of the First Serial Correlation Coefficient", Biometrika, 53, pp. 523-626.

Cox, D.R. and Lewis, P.A.W., (1966). The Statistical Analysis of Series of Events, Chapman and Hall, London, Ch. 6.

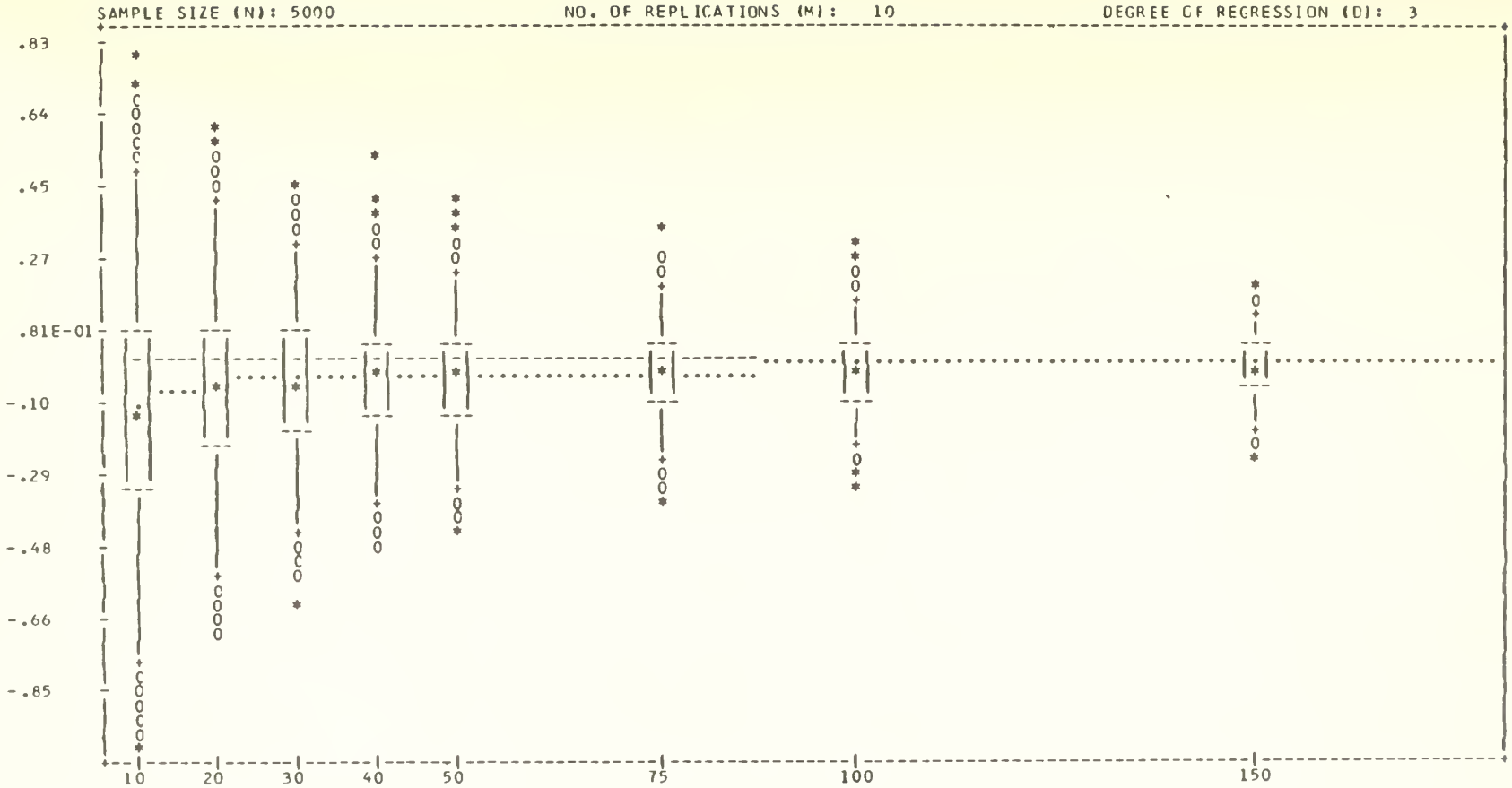
Cramer, H. (1946). Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey.

Heidelberger, P. and Lewis, P.A.W. (1981). "Regression-Adjusted Estimates for Regenerative Simulations, with Graphics", Comm. of the A.C.M., 24, pp. 260-273.

Johnson, N.L. and Kotz, S., (1970). Continuous Univariate Distributions - 1, Houghton Mifflin Company, Boston, Ch. 17.

Linnebur, D.G., (1982). "A Graphical Test Bed for Analyzing and Reporting the Results of a Simulation Experiment", Master's Thesis, Naval Postgraduate School, Monterey, California.

Quenouille, M.H., (1948). "Some Results in the Testing of Serial Correlation Coefficients", Biometrika, 35, pp. 261-267.



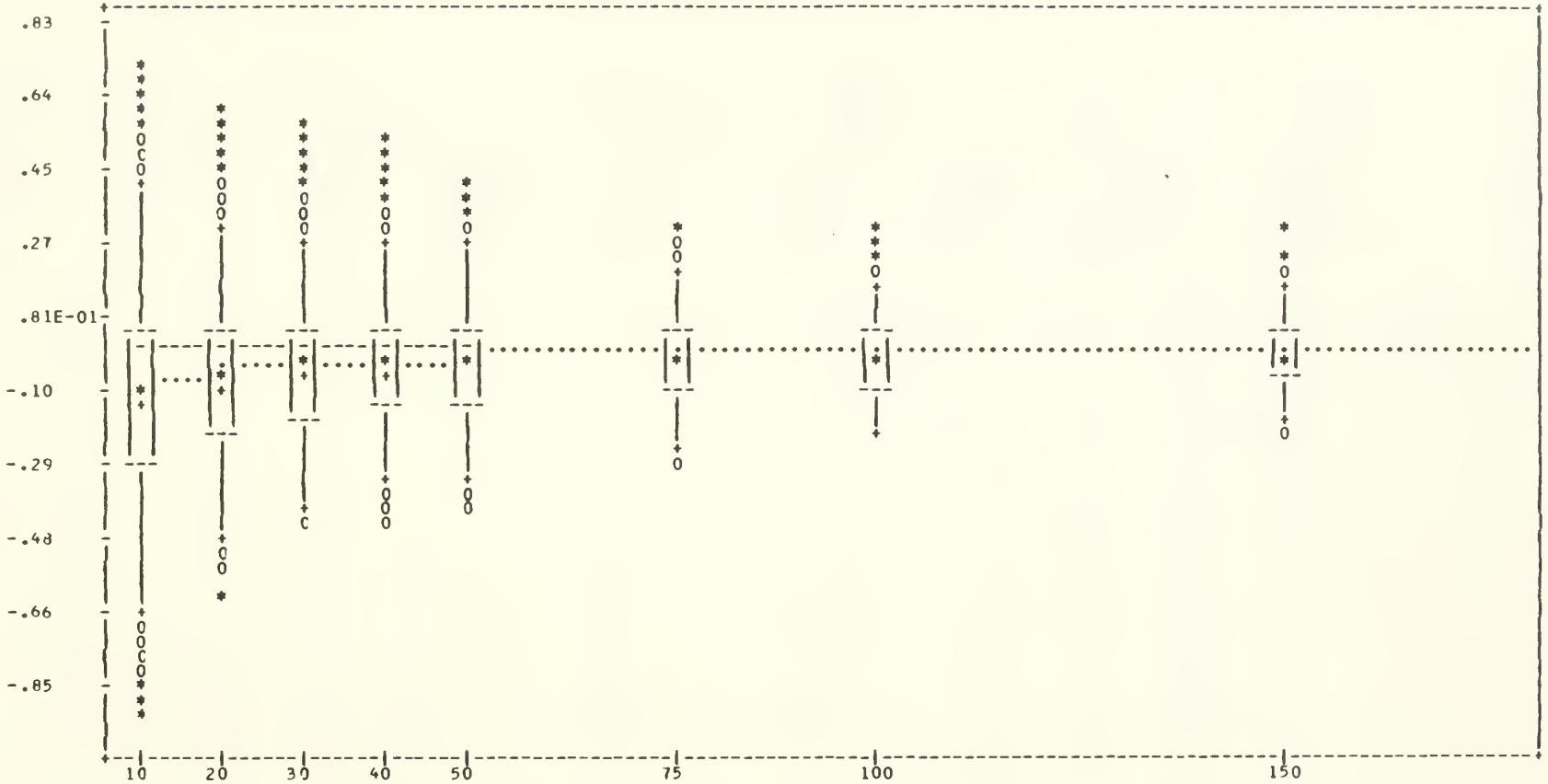
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1074	-.5125E-01	-.3465E-01	-.2611E-01	-.1980E-01	-.1442E-01	-.1008E-01	-.7372E-02
STD	.2996	.2206	.1829	.1558	.1389	.1142	.9653E-01	.7822E-01
STD MEAN	.4237E-02	.4416E-02	.4488E-02	.4405E-02	.4392E-02	.4444E-02	.4317E-02	.4306E-02
SKEWNESS	0.108	0.027	0.013	0.058	0.069	0.059	-0.066	-0.148
KURTOSIS	-0.426	-0.375	-0.274	-0.164	-0.147	-0.228	0.340	0.086

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:		-.271774E-03	-1.03074	.694455	-11.1320
VARIANCE OF REGRESSION - COEFFICIENTS:		.151458E-04	.152014E-01	7.34950	349.482
STD DEV OF REGRESSION - COEFFICIENTS:		.389176E-02	.123294	2.71100	18.6944
REGRESSION ON VARIANCE - COEFFICIENTS:		.743756	2.85748	-9.48545	6.28721

VERTICAL SCALE: YMIN = -0.9586
YMAX = 0.8258

ESTIMATOR: ESTIMATES OF THE LAG ONE SERIAL CORRELATION COEFFICIENT FOR A NORMAL(0,1) SAMPLE

Figure 1(a)



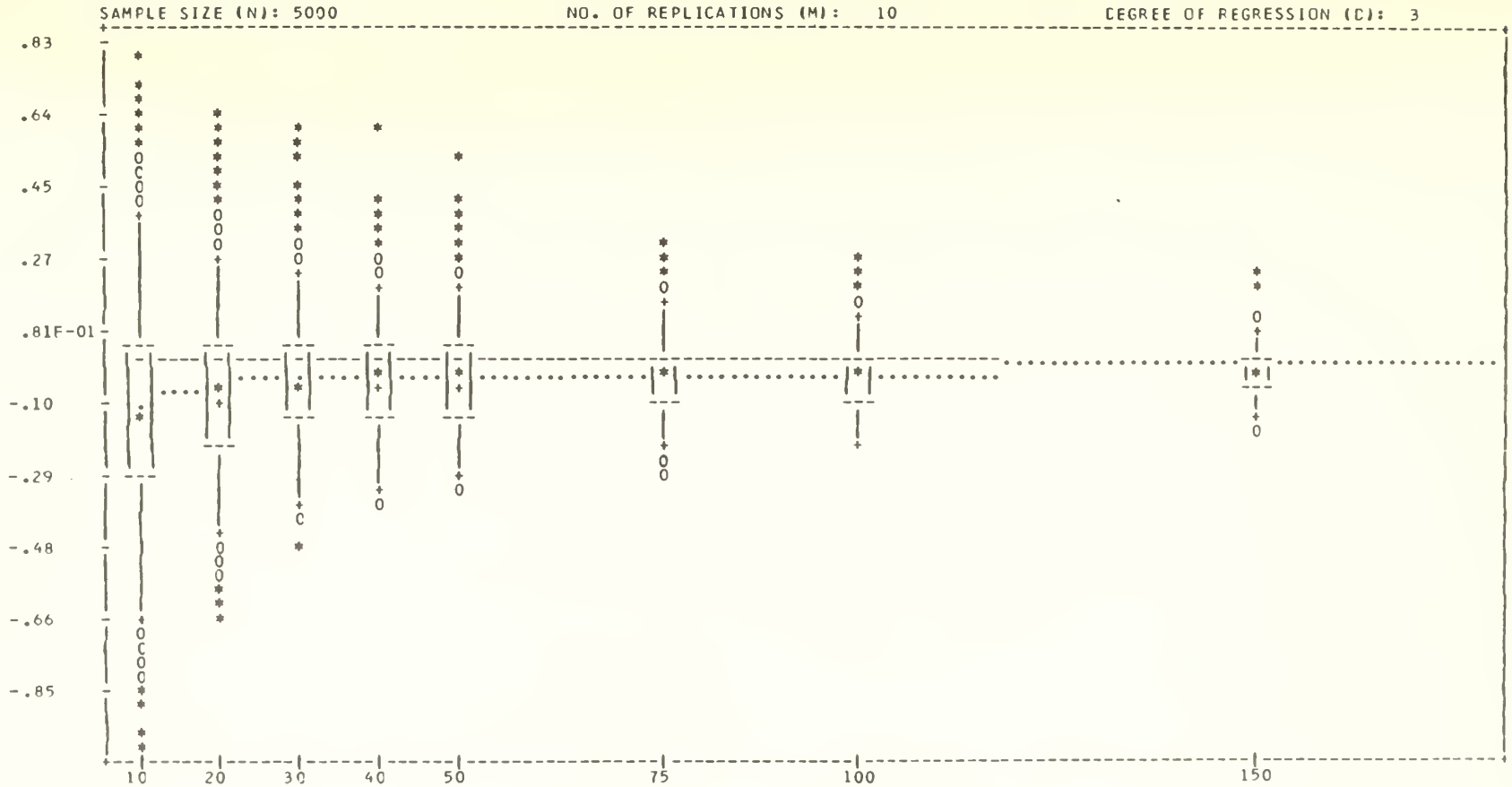
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1034	-.4659E-01	-.2814E-01	-.1817E-01	-.1248E-01	-.6363E-02	-.2612E-02	.1486E-02
STD	.2608	.1962	.1663	.1480	.1337	.1102	.9568E-01	.8356E-01
STD MEAN	.3689E-02	.3924E-02	.4080E-02	.4187E-02	.4227E-02	.4288E-02	.4275E-02	.4600E-02
SKEWNESS	0.327	0.531	0.577	0.494	0.470	0.423	0.487	0.372
KURTOSIS	-0.207	0.039	0.181	0.127	-0.049	-0.284	0.188	0.224

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.775021E-02	-.974092	-3.49581	21.0796
VARIANCE OF REGRESSION - COEFFICIENTS:	.238496E-04	.140388E-01	10.0476	575.014
STD DEV OF REGRESSION - COEFFICIENTS:	.488361E-02	.118485	3.16979	23.9794
REGRESSION ON VARIANCE - COEFFICIENTS:	1.17537	-2.52788	3.87640	-2.64231

VERTICAL SCALE: YMIN = -0.9986
YMAX = 0.8258

ESTIMATOR: ESTIMATES OF THE LAG ONE SERIAL CORRELATION COEFFICIENT FOR A CHI-SQUARE(1) SAMPLE

Figure 1(b)



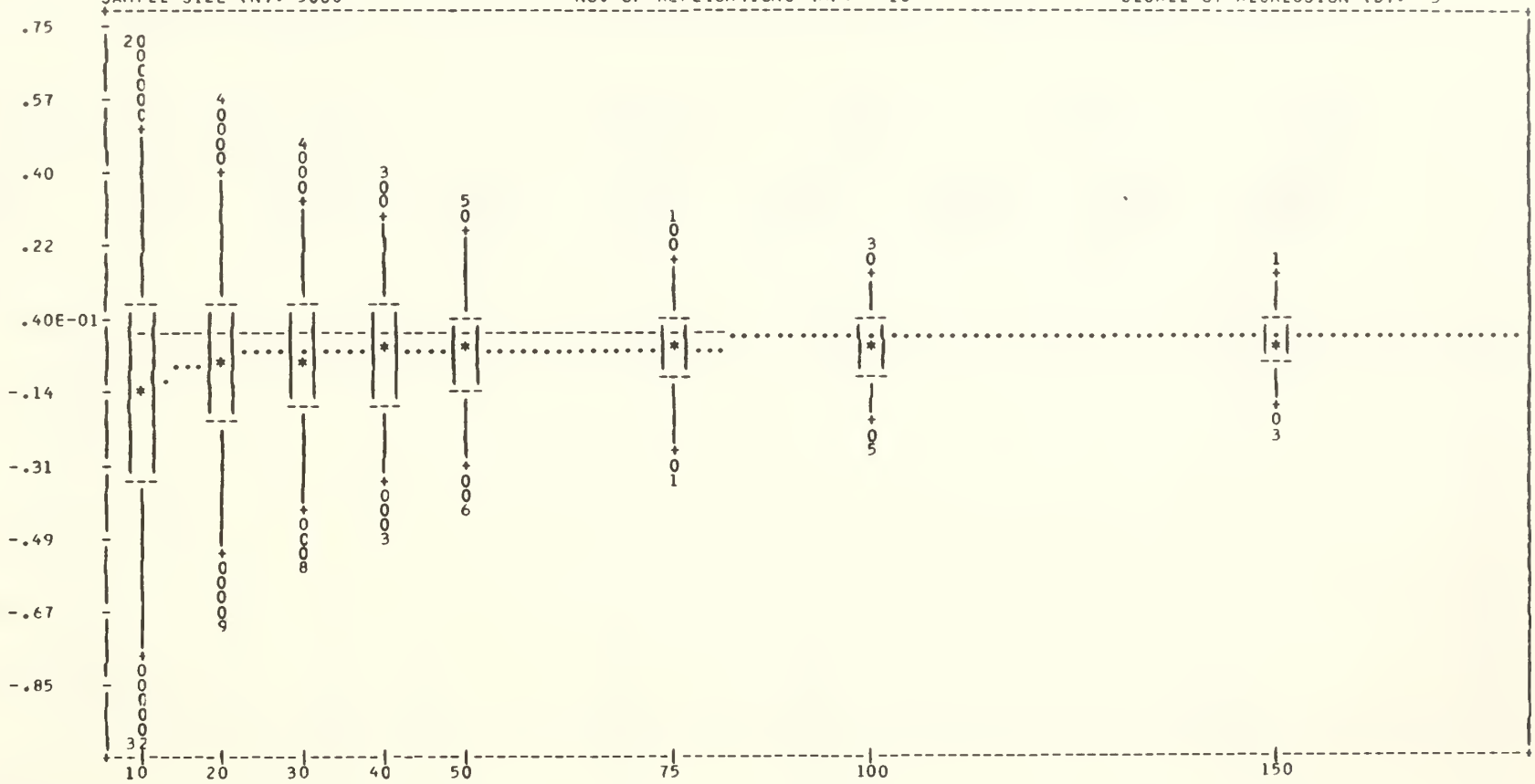
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1098	-.5318E-01	-.3701E-01	-.2851E-01	-.2317E-01	-.1698E-01	-.1271E-01	-.1049E-01
STD	.2597	.1890	.1574	.1334	.1195	.9931E-01	.8528E-01	.6686E-01
STD MEAN	.3672E-02	.3781E-02	.3864E-02	.3773E-02	.3778E-02	.3866E-02	.3814E-02	.3681E-02
SKEWNESS	0.340	0.462	0.519	0.545	0.581	0.580	0.557	0.668
KURTOSIS	-0.027	0.243	0.280	0.434	0.682	0.507	0.725	0.901

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	-.310093E-02	-1.04703	1.87581	-20.8023
VARIANCE OF REGRESSION - COEFFICIENTS:	.601527E-05	.435615E-01	22.9494	960.390
STD DEV OF REGRESSION - COEFFICIENTS:	.245260E-02	.208714	4.79055	30.9902
REGRESSION ON VARIANCE - COEFFICIENTS:	.563485	2.35455	-10.1749	12.1317

VERTICAL SCALE: YMIN = -0.9986
YMAX = 0.8258

ESTIMATOR: ESTIMATES OF THE LAG ONE SERIAL CORRELATION COEFFICIENT FOR A LOGNORMAL(0,1) SAMPLE

Figure 1(c)



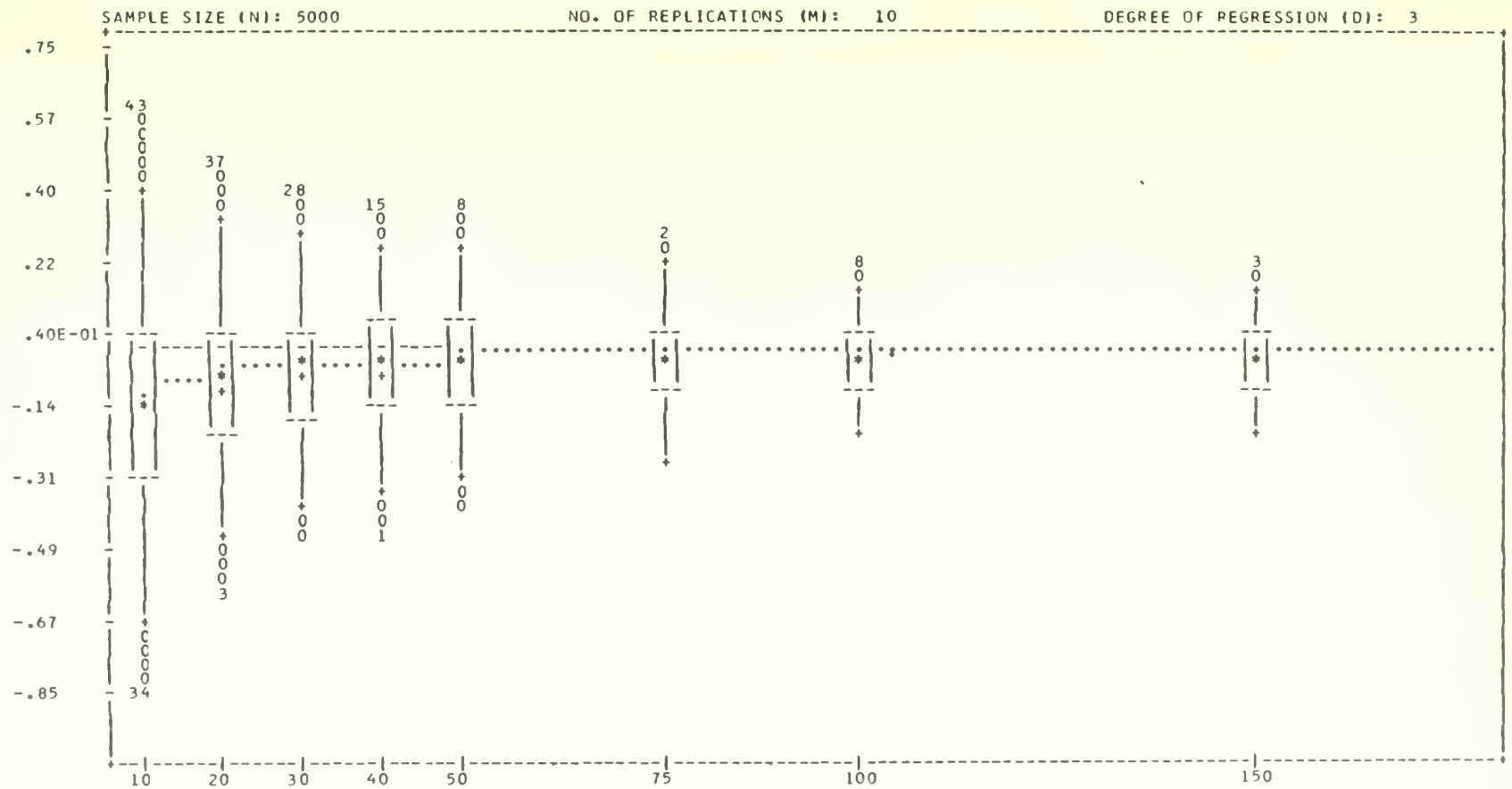
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1203	-.5408E-01	-.3592E-01	-.2671E-01	-.2014E-01	-.1459E-01	-.1015E-01	-.7443E-02
STD	.3391	.2324	.1893	.1597	.1416	.1156	.9758E-01	.7874E-01
STD MEAN	.4796E-02	.4648E-02	.4646E-02	.4517E-02	.4479E-02	.4500E-02	.4364E-02	.4334E-02
SKEWNESS	-0.346	-0.044	-0.034	0.045	0.059	0.052	-0.077	-0.156
KURTOSIS	2.547	-0.074	-0.056	0.015	-0.014	-0.155	0.448	0.144

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	-.337710E-03	-1.02174	-.542526	-12.4196
VARIANCE OF REGRESSION - COEFFICIENTS:	.149923E-04	.141879E-01	7.25536	364.315
STD DEV OF REGRESSION - COEFFICIENTS:	.387199E-02	.119113	2.69358	19.0871
REGRESSION ON VARIANCE - COEFFICIENTS:	.713406	3.47347	-11.6751	15.9883

VERTICAL SCALE: YMIN = -0.9890
YMAX = 0.7503

ESTIMATOR: ESTIMATES OF THE Z-TRANSFORM OF THE SERIAL CORRELATION COEFFICIENT FOR A NORMAL(0,1) SAMPLE

Figure 2(a)



SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1103	-.4744E-01	-.2797E-01	-.1800E-01	-.1231E-01	-.6246E-02	-.2488E-02	.1572E-02
STD	.2832	.2043	.1713	.1515	.1361	.1114	.9663E-01	.8421E-01
STD MEAN	.4005E-02	.4085E-02	.4204E-02	.4286E-02	.4303E-02	.4335E-02	.4321E-02	.4635E-02
SKEWNESS	0.146	0.561	0.631	0.534	0.497	0.435	0.513	0.394
KURTOSIS	0.474	0.327	0.470	0.362	0.100	-0.221	0.295	0.323

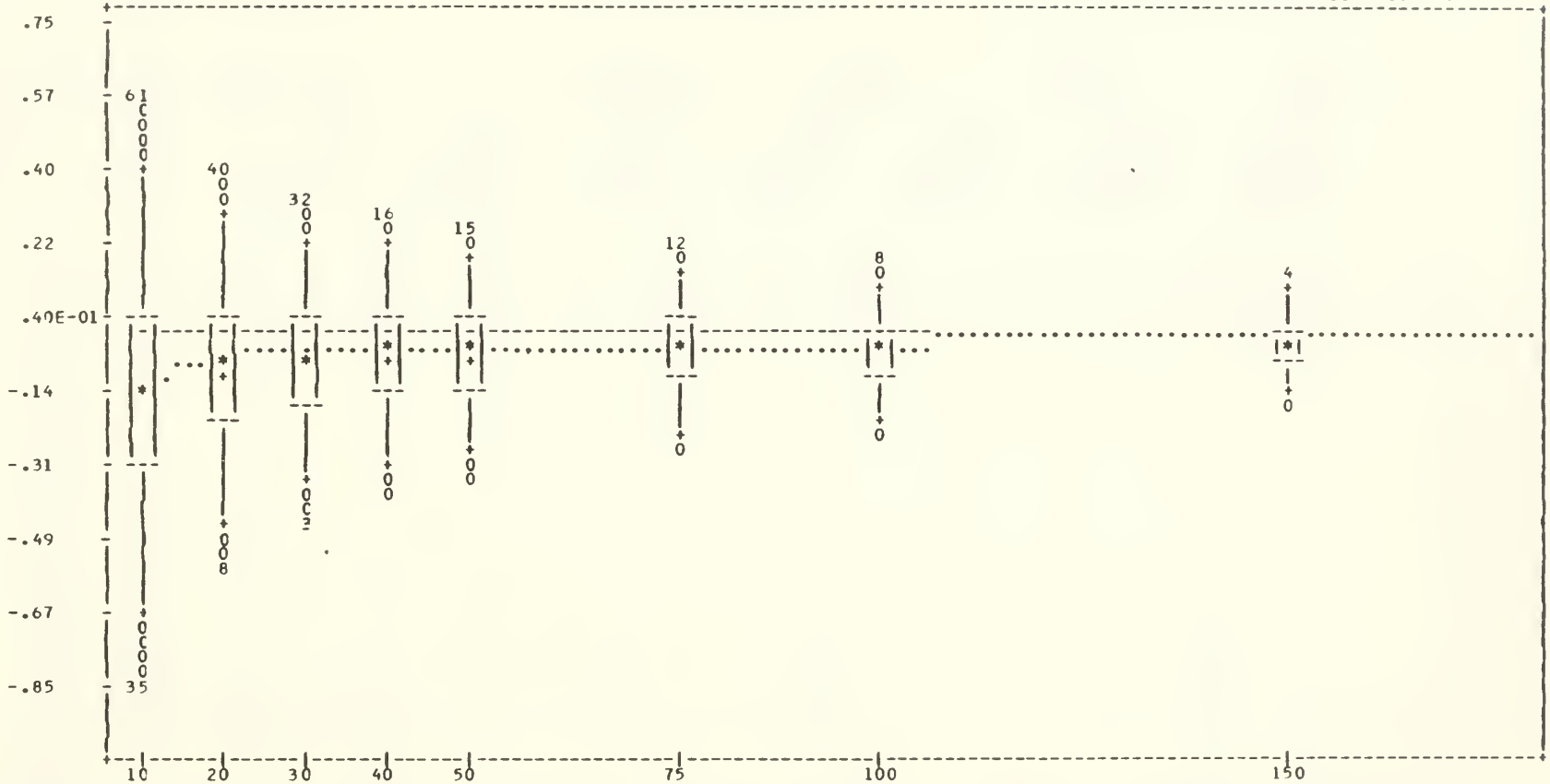
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.770474E-02	-.955687	-3.68524	14.3363
VARIANCE OF REGRESSION - COEFFICIENTS:	.240493E-04	.138909E-01	10.6408	618.694
STD DEV OF REGRESSION - COEFFICIENTS:	.490401E-02	.117860	3.26203	24.8736
REGRESSION ON VARIANCE - COEFFICIENTS:	1.14227	-1.82462	1.51063	2.69722

VERTICAL SCALE: YMIN = -0.9890
YMAX = 0.7503

ESTIMATOR: ESTIMATES OF THE Z-TRANSFORM OF THE SERIAL CORRELATION COEFFICIENT FOR A CHI-SQUARE(1) SAMPLE

Figure 2(b)

SAMPLE SIZE (N): 5000 NO. OF REPLICATIONS (M): 10 DEGREE OF REGRESSION (D): 3

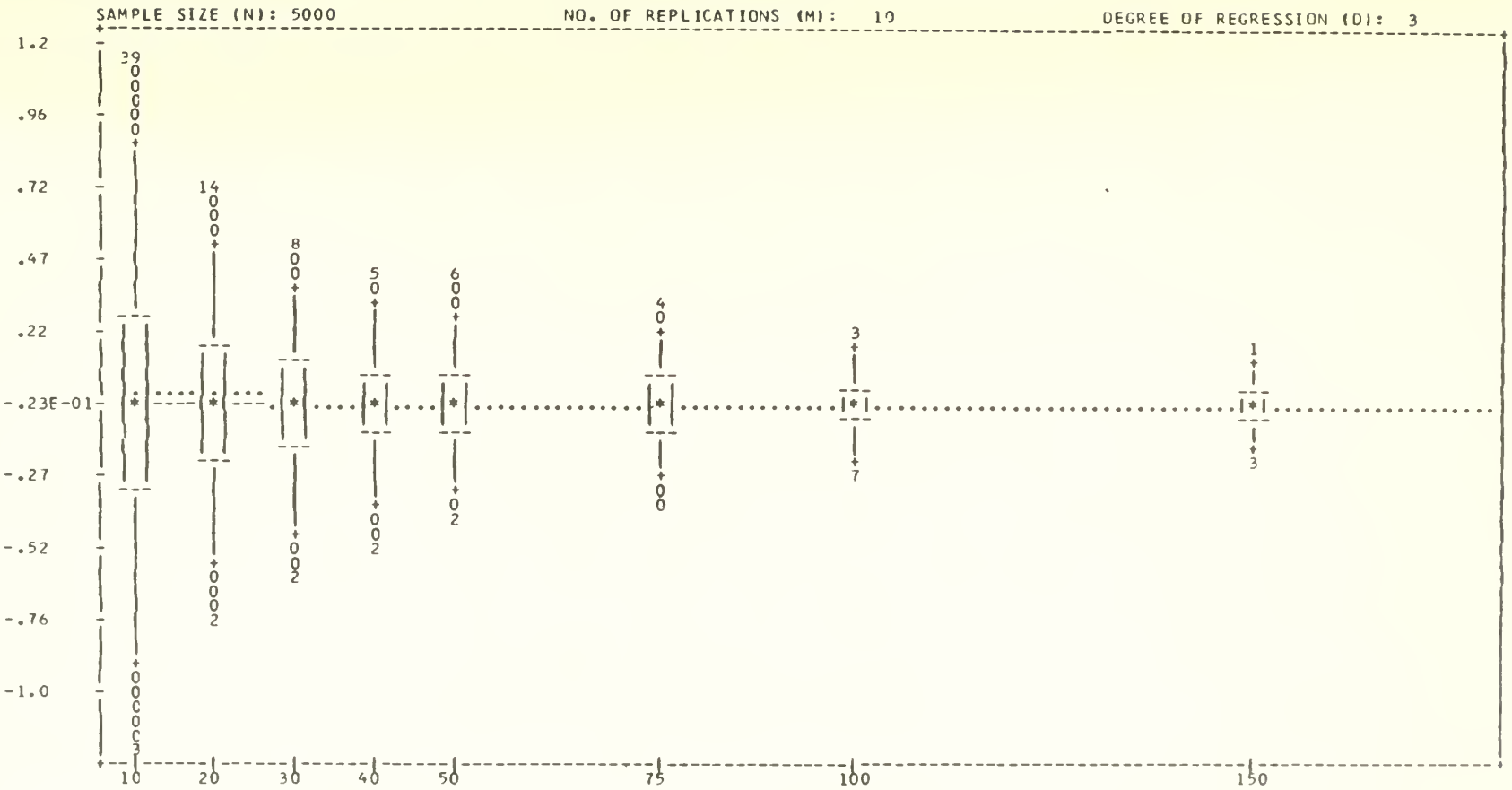


SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	-.1172	-.5411E-01	-.3724E-01	-.2857E-01	-.2316E-01	-.1695E-01	-.1268E-01	-.1046E-01
STD	.2849	.1970	.1619	.1362	.1216	.1004	.8603E-01	.6723E-01
STD MEAN	.4029E-02	.3941E-02	.3973E-02	.3852E-02	.3846E-02	.3909E-02	.3847E-02	.3701E-02
SKEWNESS	0.969	0.474	0.556	0.594	0.632	0.600	0.616	0.685
KURTOSIS	1.527	0.656	0.611	0.819	1.032	0.631	0.835	0.995

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:		-.312200E-02	-1.03891	1.64581	-26.7689
VARIANCE OF REGRESSION - COEFFICIENTS:		.589428E-05	.425660E-01	22.7239	967.803
STD DEV OF REGRESSION - COEFFICIENTS:		.242782E-02	.206315	4.76696	31.1095
REGRESSION ON VARIANCE - COEFFICIENTS:		.545919	2.71325	-11.2129	16.7234

VERTICAL SCALE: YMIN = -0.9890
YMAX = 0.7503

ESTIMATOR: ESTIMATES OF THE Z-TRANSFORM OF THE SERIAL CORRELATION COEFFICIENT FOR A LOGNORMAL(0,1) SAMPLE



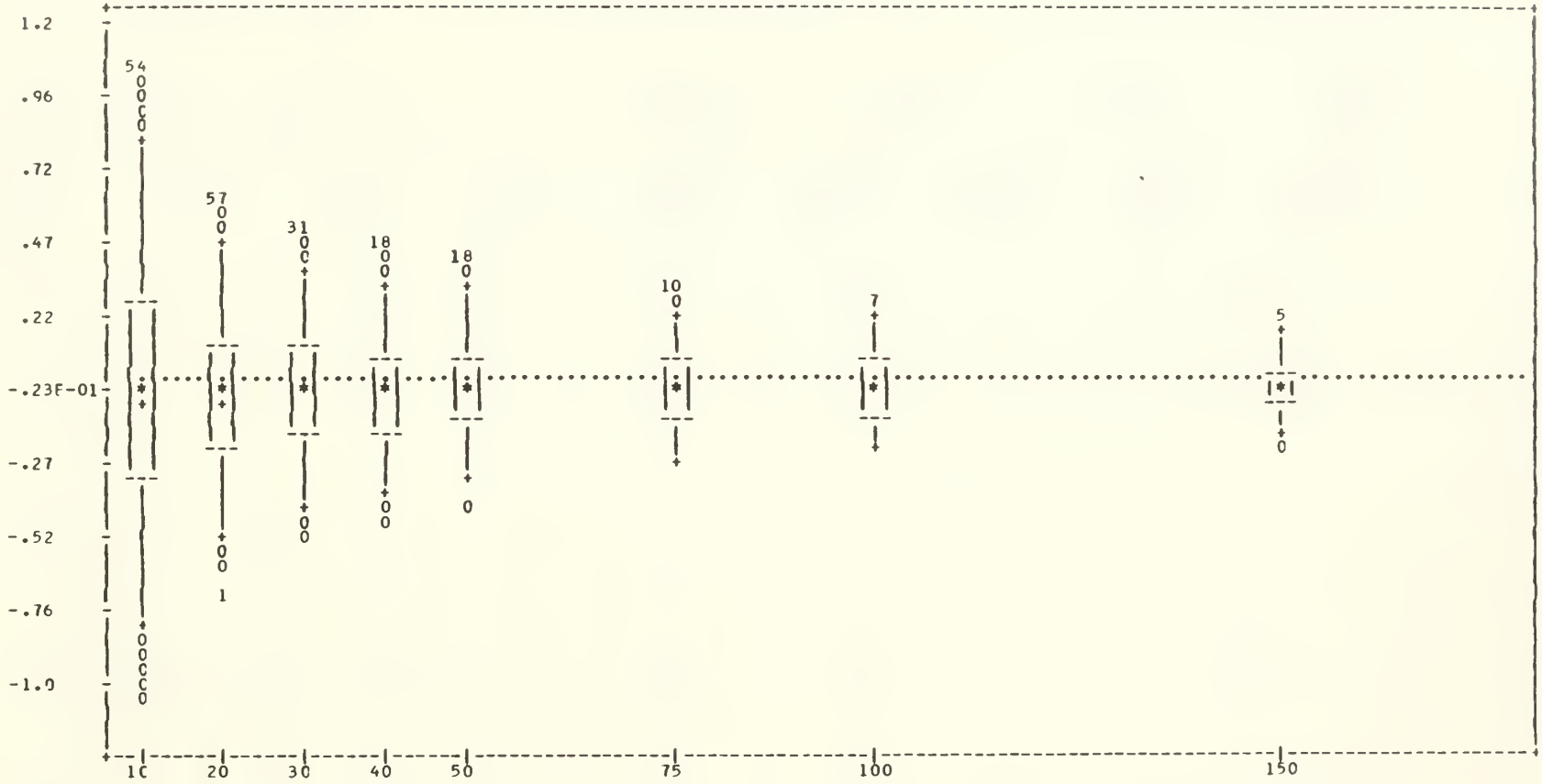
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	.1471E-01	.4979E-02	.3881E-04	-.9342E-03	.2980E-02	-.9550E-03	-.3521E-03	-.3150E-03
STD	.4475	.2757	.2137	.1733	.1543	.1257	.1028	.8101E-01
STD MEAN	.6328E-02	.3513E-02	.5245E-02	.4901E-02	.4879E-02	.4892E-02	.4556E-02	.4459E-02
SKEWNESS	0.319	0.190	0.125	0.130	0.152	0.134	-0.017	-0.126
KURTOSIS	0.094	0.008	-0.125	-0.031	0.059	-0.106	0.340	0.045

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.546557E-03	-.131495	5.60890	-28.8484
VARIANCE OF REGRESSION - COEFFICIENTS:	.153676E-04	.906536E-01	89.6084	5054.53
STD DEV OF REGRESSION - COEFFICIENTS:	.392016E-02	.301087	9.46617	71.0952
REGRESSION ON VARIANCE - COEFFICIENTS:	.864258	1.28888	7.48415	-.560686

VERTICAL SCALE: YMIN = -1.2072
YMAX = 1.2114

ESTIMATOR: ESTIMATES OF THE 2-FOLD JACKNIFED SERIAL CORRELATION COEFFICIENT FOR A NORMAL(0,1) SAMPLE

Figure 3(a)



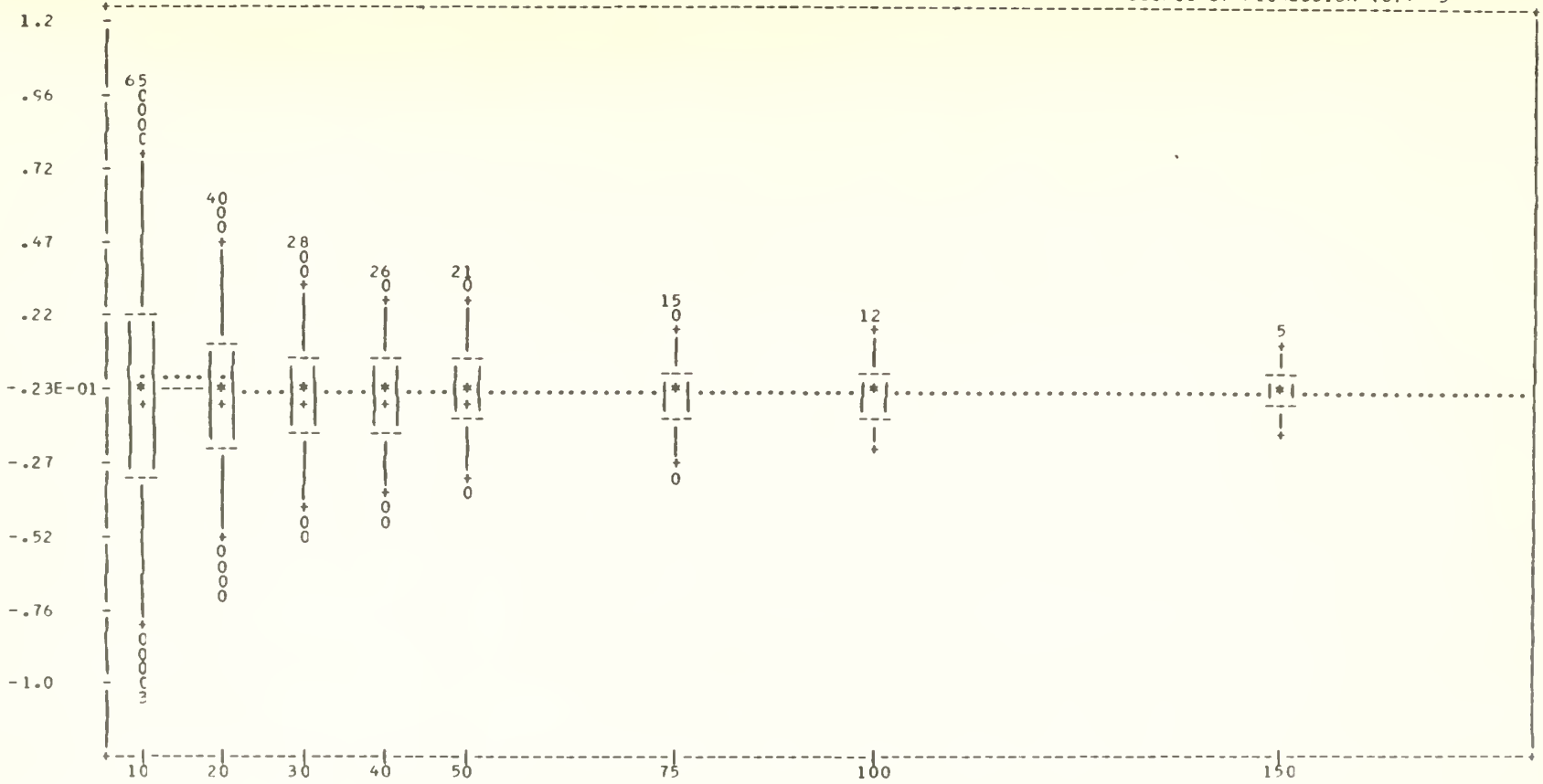
SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	.2278E-01	.9513E-02	.1008E-01	.1067E-01	.1150E-01	.8868E-02	.7255E-02	.9341E-02
STD	.4095	.2680	.2152	.1824	.1628	.1324	.1119	.9194E-01
STD MEAN	.5792E-02	.5361E-02	.5282E-02	.5159E-02	.5149E-02	.5154E-02	.5062E-02	.5061E-02
SKEWNESS	0.544	0.804	0.813	0.683	0.694	0.724	0.585	0.581
KURTOSIS	0.138	0.701	0.888	0.424	0.444	0.446	0.187	0.665

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.578116E-02	.412461	-11.0824	86.4791
VARIANCE OF REGRESSION - COEFFICIENTS:	.270782E-04	.844226E-01	98.9634	5462.96
STD DEV OF REGRESSION - COEFFICIENTS:	.520367E-02	.290556	9.94501	73.9118
REGRESSION ON VARIANCE - COEFFICIENTS:	1.11376	2.29230	-8.84446	22.8625

VERTICAL SCALE: YMIN = -1.2072
YMAX = 1.2114

ESTIMATOR: ESTIMATES OF THE 2-FOLD JACKNIFED SERIAL CORRELATION COEFFICIENT FOR A CHI-SQUARE(1) SAMPLE

Figure 3(b)



SUBSAMPLE SIZE	10	20	30	40	50	75	100	150
MEAN	.1258E-01	.3435E-02	-.3343E-02	-.3821E-02	-.1695E-02	-.5042E-02	-.2255E-02	-.3986E-02
STD	.4028	.2600	.2011	.1724	.1533	.1222	.1011	.7999E-01
STD MEAN	.5696E-02	.5200E-02	.4936E-02	.4876E-02	.4849E-02	.4759E-02	.4519E-02	.4403E-02
SKEWNESS	0.550	0.712	0.741	0.831	0.894	0.819	0.744	0.850
KURTOSIS	0.352	0.662	0.905	1.077	1.602	1.150	0.629	1.451

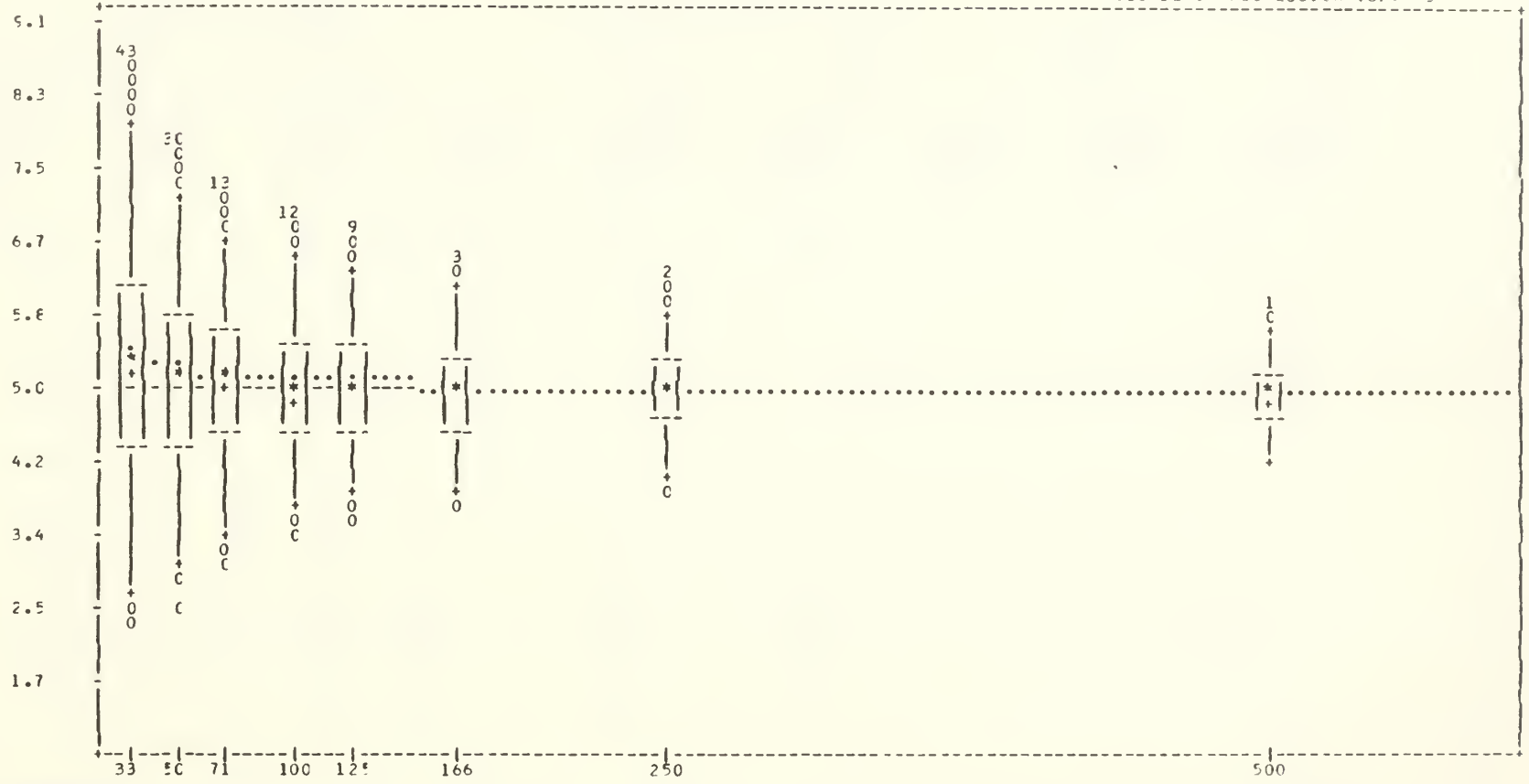
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	-.134541E-02	-.337504	12.0957	-73.3541
VARIANCE OF REGRESSION - COEFFICIENTS:	.167209E-04	.218834	154.825	7199.73
STD DEV OF REGRESSION - COEFFICIENTS:	.408512E-02	.467757	12.4429	84.8512
REGRESSION ON VARIANCE - COEFFICIENTS:	.617429	5.83375	-19.4923	35.1018

VERTICAL SCALE: YMIN = -1.2072
YMAX = 1.2114

ESTIMATOR: ESTIMATES OF THE 2-FOLD JACKNIFED SERIAL CORRELATION COEFFICIENT FOR A LOGNORMAL(0,1) SAMPLE

Figure 3(a)

SAMPLE SIZE (N): 2500 NO. OF REPLICATIONS (M): 20 DEGREE OF REGRESSION (D): 3



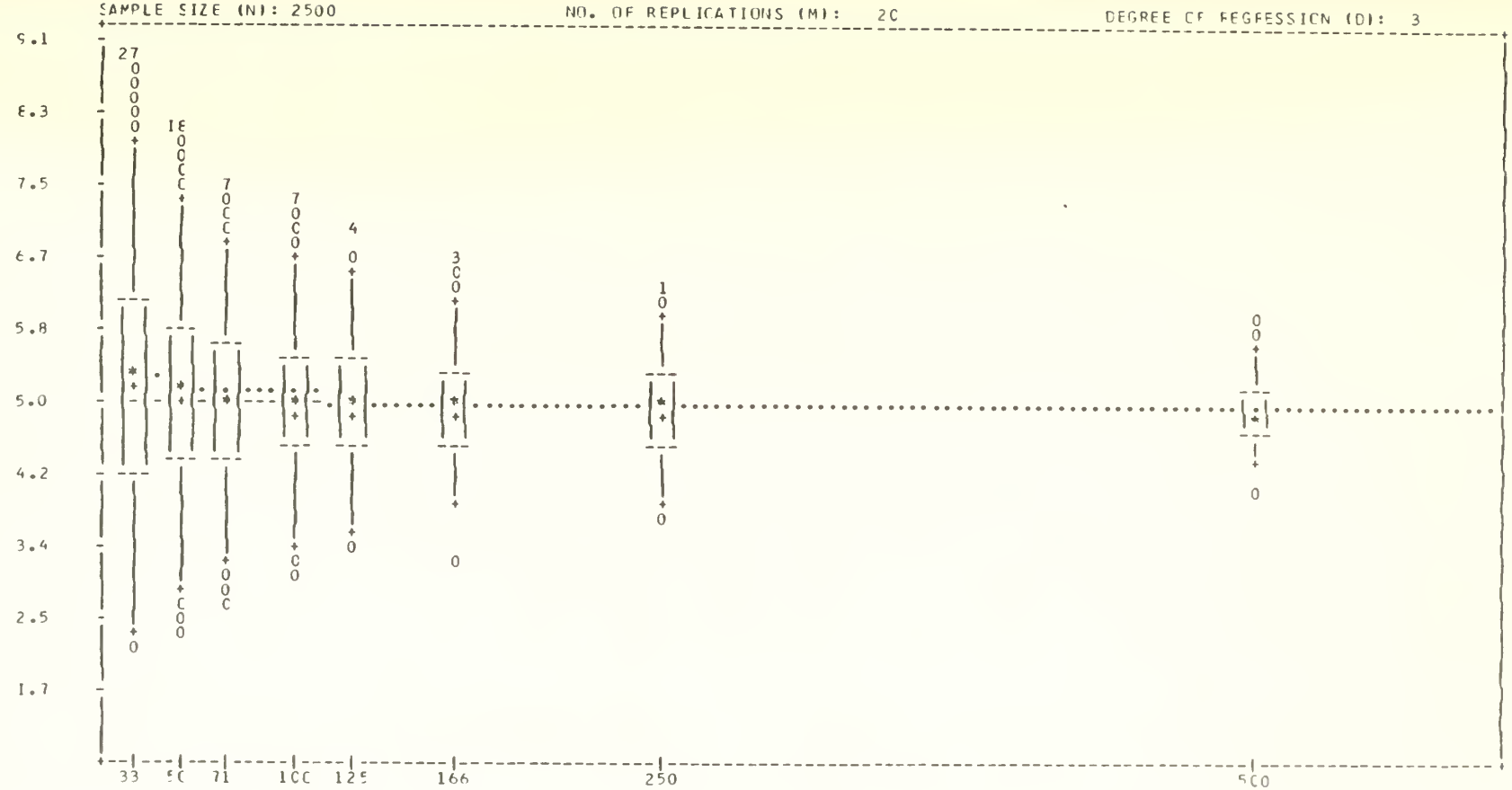
SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	5.477	5.309	5.202	5.145	5.118	5.085	5.045	5.026
STD	1.448	1.137	.8725	.7671	.6650	.5622	.4807	.3654
STD MEAN	.3738E-01	.3595E-01	.3298E-01	.3431E-01	.3325E-01	.3246E-01	.3395E-01	.3654E-01
SKEWNESS	1.217	1.074	0.732	0.812	0.680	0.525	0.354	0.679
KURTOSIS	2.872	2.198	0.904	0.762	0.693	0.157	0.246	0.119

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	4.99635	13.5123	140.737	-2063.17
VARIANCE OF REGRESSION - COEFFICIENTS:	.110506E-02	26.9280	158094.	.579950E+08
STD DEV OF REGRESSION - COEFFICIENTS:	.332424E-01	6.07684	397.610	7615.39
REGRESSION ON VARIANCE - COEFFICIENTS:	116.458	-1681.96	14121.1	-34565.3

VERTICAL SCALE: YMIN = 1.0544
YMAX = 9.1332

ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=5.

Figure 4(a)



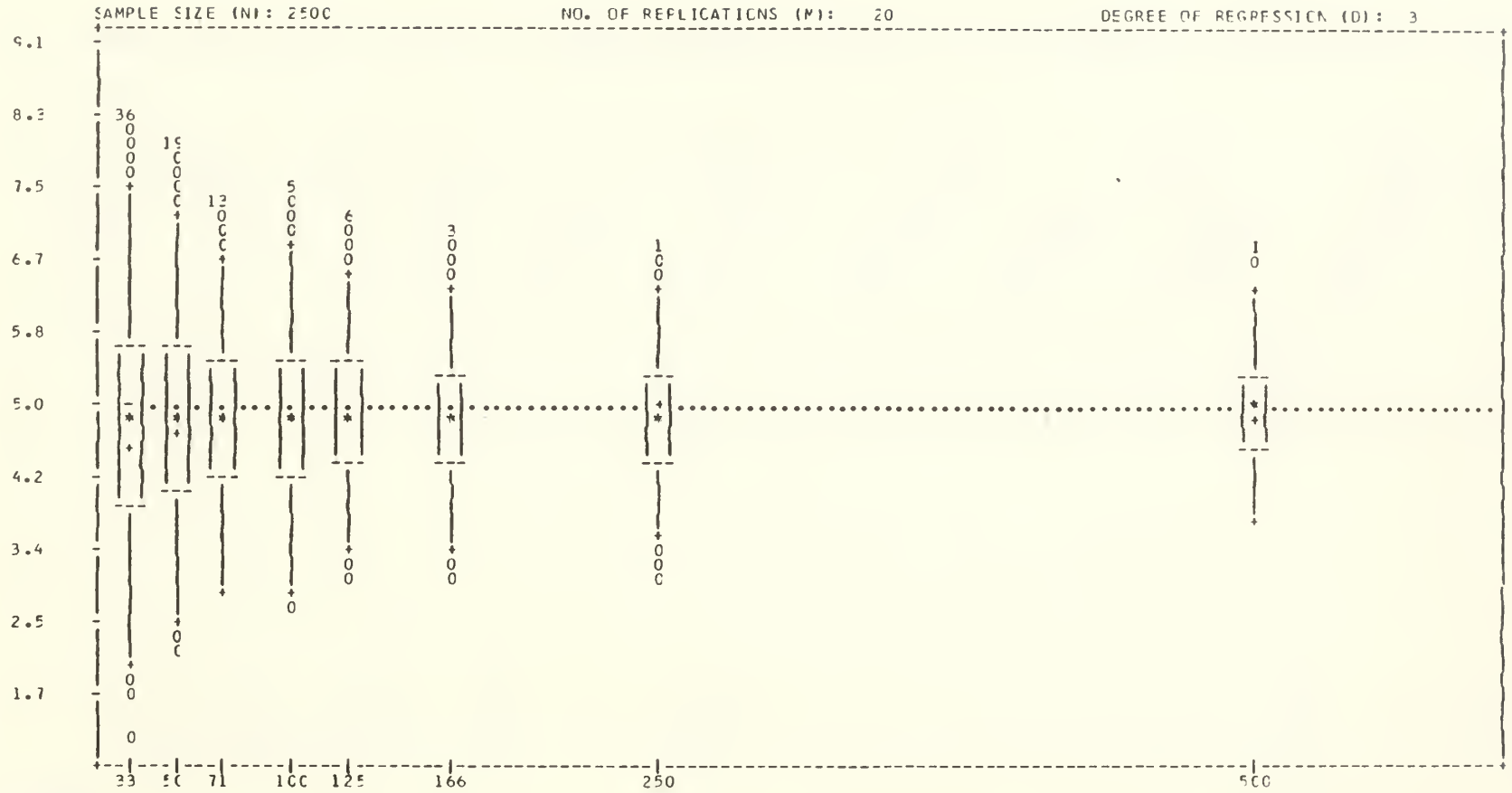
SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	5.371	5.234	5.151	5.113	5.080	5.054	5.028	4.997
STD	1.482	1.171	.9606	.8281	.7313	.6277	.5268	.3750
STD MEAN	.3827E-01	.3721E-01	.3631E-01	.3703E-01	.3667E-01	.3624E-01	.3725E-01	.3760E-01
SKEWNESS	0.858	0.702	0.591	0.572	0.525	0.429	0.354	0.482
KURTOSIS	1.572	1.009	0.670	0.375	0.373	-0.184	-0.120	-0.135

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	4.96799	15.4346	-196.572	4202.59
VARIANCE OF REGRESSION - COEFFICIENTS:	.966283E-03	4.14804	23214.0	.989117E+07
STD DEV OF REGRESSION - COEFFICIENTS:	.310851E-01	2.03667	152.361	3145.01
REGRESSION ON VARIANCE - COEFFICIENTS:	81.7175	-336.834	2162.00	-3040.92

VERTICAL SCALE: YMIN = 1.0544
YMAX = 9.1332

ESTIMATOR: MOMENT ESTIMATOR (RECIPROCAL OF SQUARED COEFFICIENT OF VARIATION) OF THE SHAPE PARAMETER OF GAMMA DISTRIBUTION K=5.

Figure 4(b)

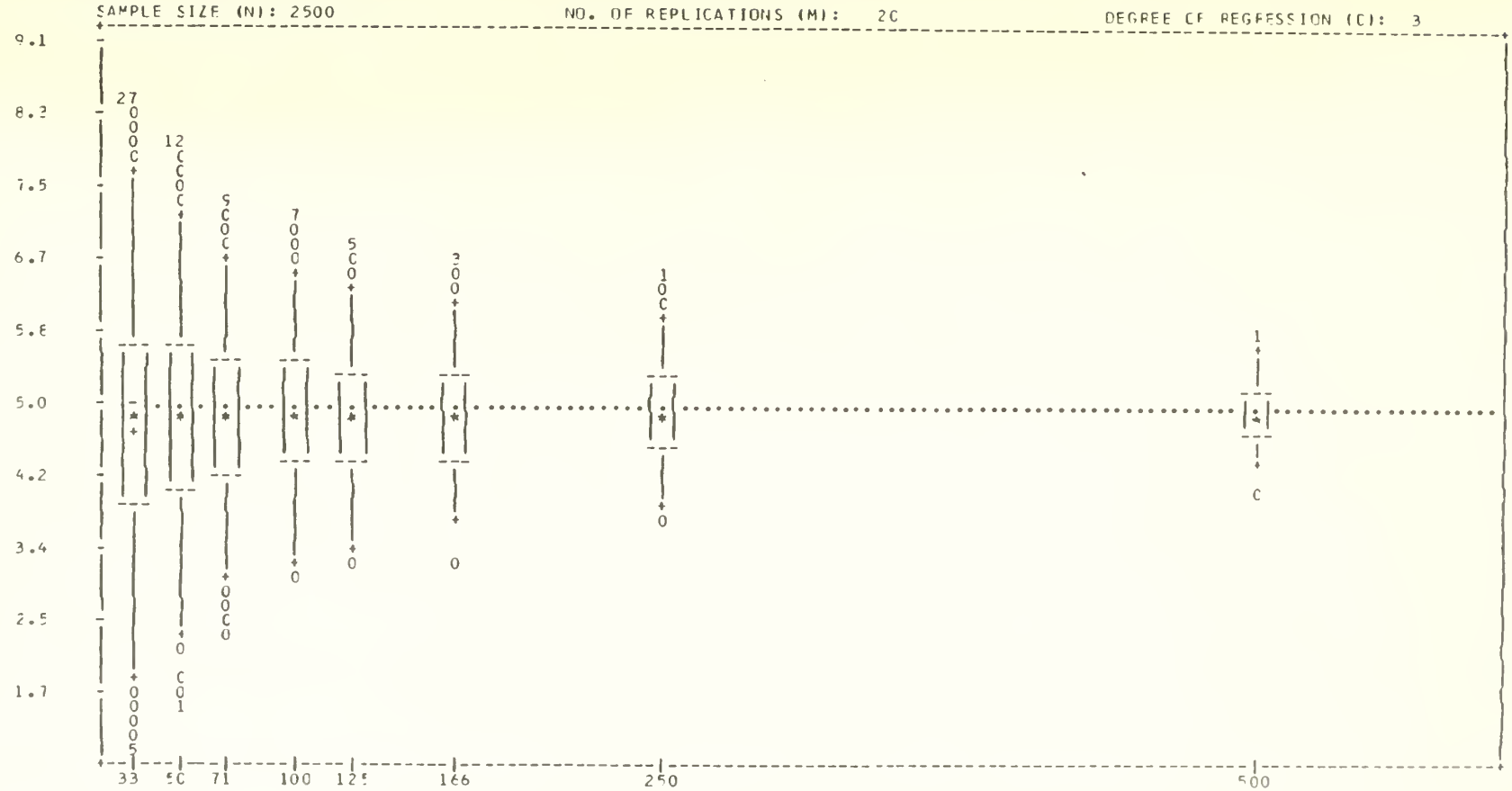


SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	4.917	4.964	4.970	4.977	5.008	5.010	4.968	5.031
STD	1.520	1.241	1.012	.9542	.8517	.7624	.7210	.6894
STD MEAN	.3925E-01	.3923E-01	.3826E-01	.4267E-01	.4259E-01	.4402E-01	.5098E-01	.6894E-01
SKEWNESS	1.177	0.960	0.912	0.616	0.524	0.364	-0.049	0.465
KURTOSIS	2.840	1.823	1.785	0.230	0.361	-0.155	0.015	-0.244

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	5.02893	-7.19724	349.841	-7711.72
VARIANCE OF REGRESSION - COEFFICIENTS:	.103702E-01	550.288	.257296E+07	.937448E+09
STD DEV OF REGRESSION - COEFFICIENTS:	.101834	24.2958	1604.05	30617.6
REGRESSION ON VARIANCE - COEFFICIENTS:	587.334	-11703.4	87638.4	-214209.

VERTICAL SCALE: YMIN = 1.0544
YMAX = 9.1332

ESTIMATOR: 4-FOLD JACKNIFED MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=5.



SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	4.919	4.951	4.958	4.984	4.977	4.980	4.979	4.969
STD	1.530	1.218	.9855	.8476	.7416	.6342	.5298	.3734
STD MEAN	.3949E-01	.3853E-01	.3725E-01	.3791E-01	.3708E-01	.3662E-01	.3746E-01	.3734E-01
SKEWNESS	0.651	0.464	0.524	0.517	0.519	0.437	0.325	0.441
KURTOSIS	1.569	0.510	0.738	0.376	0.376	-0.162	-0.132	-0.166

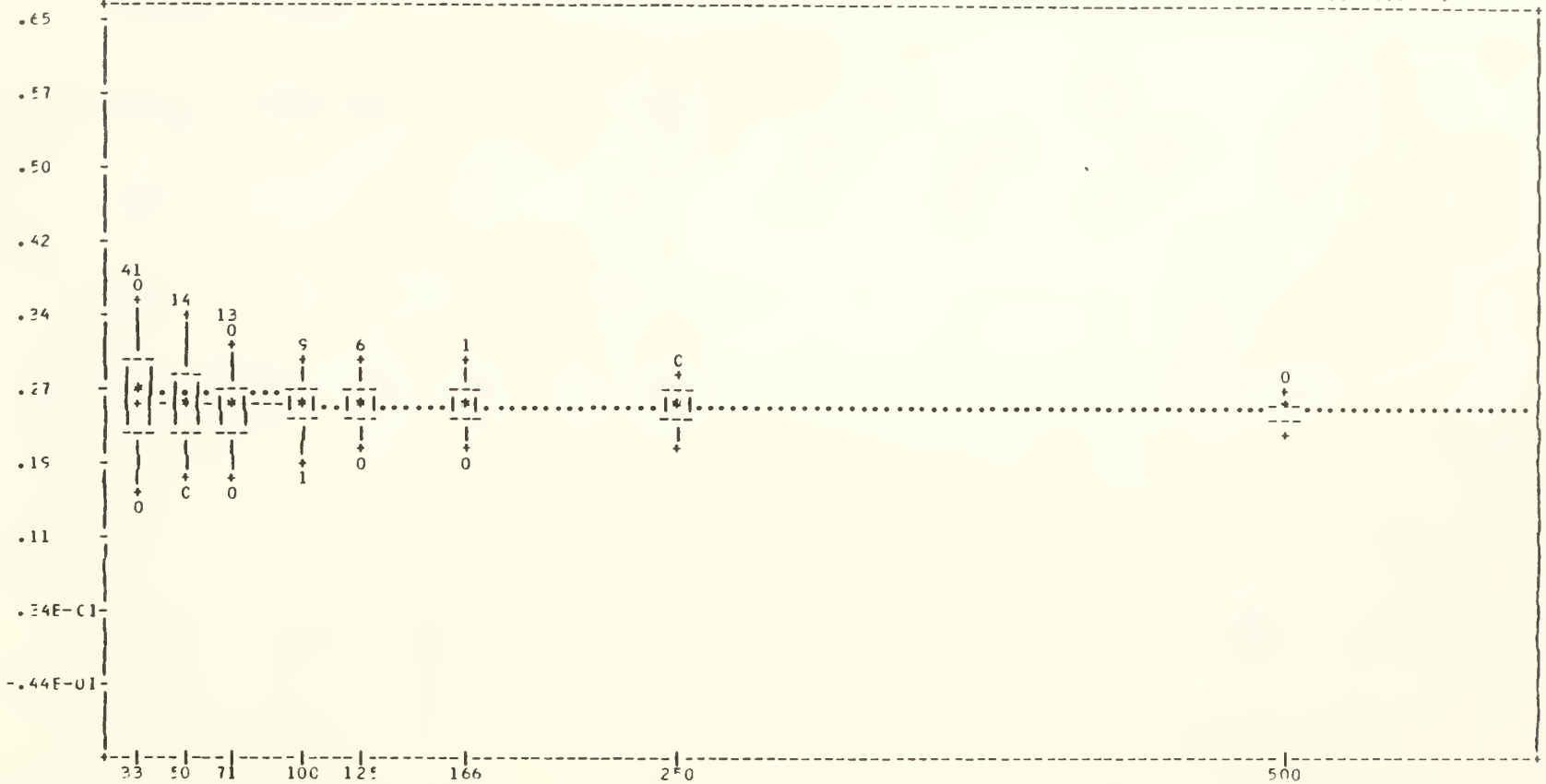
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	4.96213	5.41940	-472.241	8152.19
VARIANCE OF REGRESSION - COEFFICIENTS:	.905102E-03	13.9244	91080.1	.426571E+08
STD DEV OF REGRESSION - COEFFICIENTS:	.301513E-01	3.73153	303.117	6531.21
REGRESSION ON VARIANCE - COEFFICIENTS:	81.3402	-382.332	3397.67	-7671.09

VERTICAL SCALE: YMIN = 1.0544
YMAX = 9.1332

ESTIMATOR: 4-FOLD JACKKNIFED MOMINT ESTIMATOR OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=5.

Figure 4(d)

SAMPLE SIZE (N): 2500 NO. OF REPLICATIONS (M): 20 DEGREE OF REGRESSION (D): 3

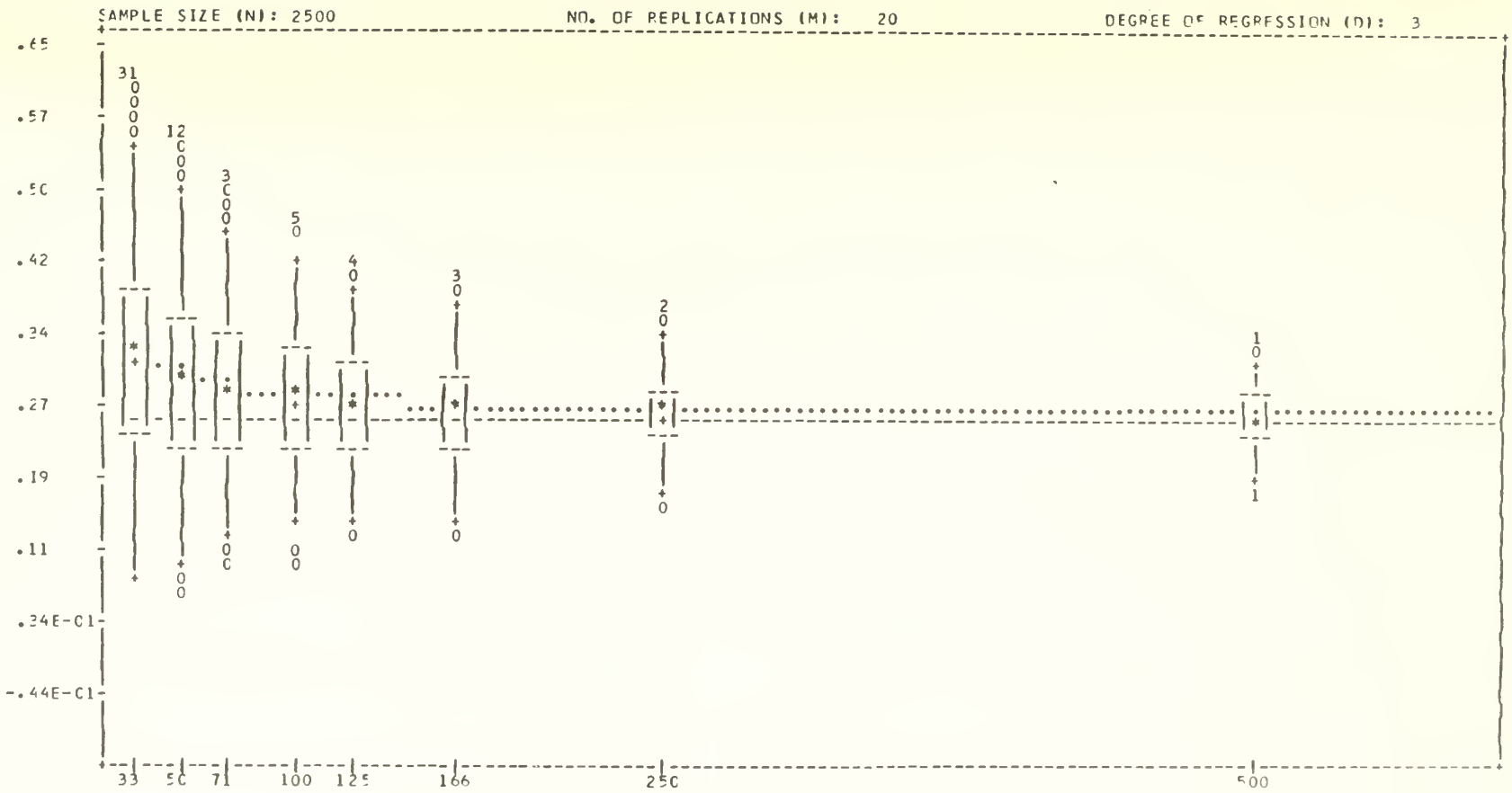


SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.2663	.2625	.2596	.2570	.2561	.2554	.2542	.2533
STD	.5589E-01	.4315E-01	.3680E-01	.2865E-01	.2595E-01	.2304E-01	.1821E-01	.1255E-01
STD MEAN	.1443E-02	.1366E-02	.1391E-02	.1281E-02	.1298E-02	.1330E-02	.1288E-02	.1255E-02
SKEWNESS	1.015	0.750	0.741	0.408	0.448	0.326	0.171	0.017
KURTOSIS	1.771	0.950	0.709	-0.035	0.235	0.170	-0.301	-0.120

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.252401	.445971	4.43231	-60.8146
VARIANCE OF REGRESSION - COEFFICIENTS:	.198013E-05	.662472E-02	40.5503	18732.8
STD DEV OF REGRESSION - COEFFICIENTS:	.140717E-02	.813924E-01	6.36791	136.868
REGRESSION ON VARIANCE - COEFFICIENTS:	.733840E-01	.161415	-.519010	3.25438

VERTICAL SCALE: YMIN = -0.1055
YMAX = 0.6520

ESTIMATOR: MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=.25



SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.3288	.3064	.2912	.2814	.2759	.2712	.2658	.2600
STD	.1209	.9825E-01	.8061E-01	.7022E-01	.6193E-01	.5624E-01	.4595E-01	.3546E-01
STD MEAN	.2123E-02	.3107E-02	.3047E-02	.3140E-02	.3096E-02	.3247E-02	.3252E-02	.3546E-02
SKEWNESS	0.765	0.565	0.345	0.399	0.351	0.254	0.355	0.362
KURTOSIS	0.746	0.347	-0.117	0.467	0.206	-0.106	0.240	0.728

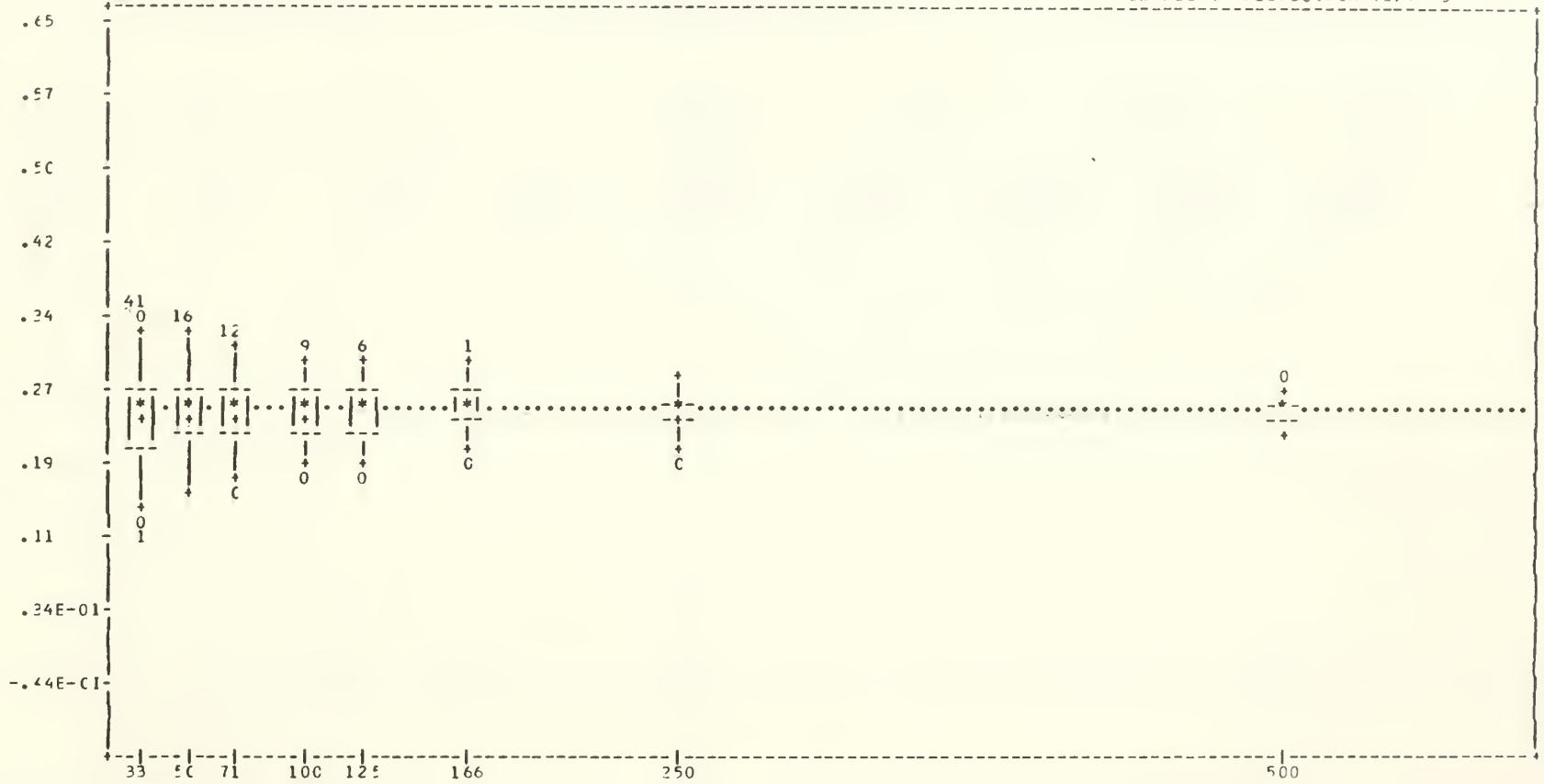
MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.254517	2.67586	-1.86707	-122.781
VARIANCE OF REGRESSION - COEFFICIENTS:	.970740E-05	.211753	874.454	364349.
STD DEV OF REGRESSION - COEFFICIENTS:	.311567E-02	.460167	29.5712	603.614
REGRESSION ON VARIANCE - COEFFICIENTS:	.974941	-11.0911	80.0564	-187.121

VERTICAL SCALE: YMIN = -0.1055
YMAX = 0.6520

ESTIMATOR: MOMENT ESTIMATOR (RECIPROCAL OF SQUARED COEFFICIENT OF VARIATION) OF THE SHAPE PARAMETER OF GAMMA DISTRIBUTION K=3

Figure 5(b)

SAMPLE SIZE (N): 2500 NO. OF REPLICATIONS (M): 20 DEGREE OF REGRESSION (D): 3



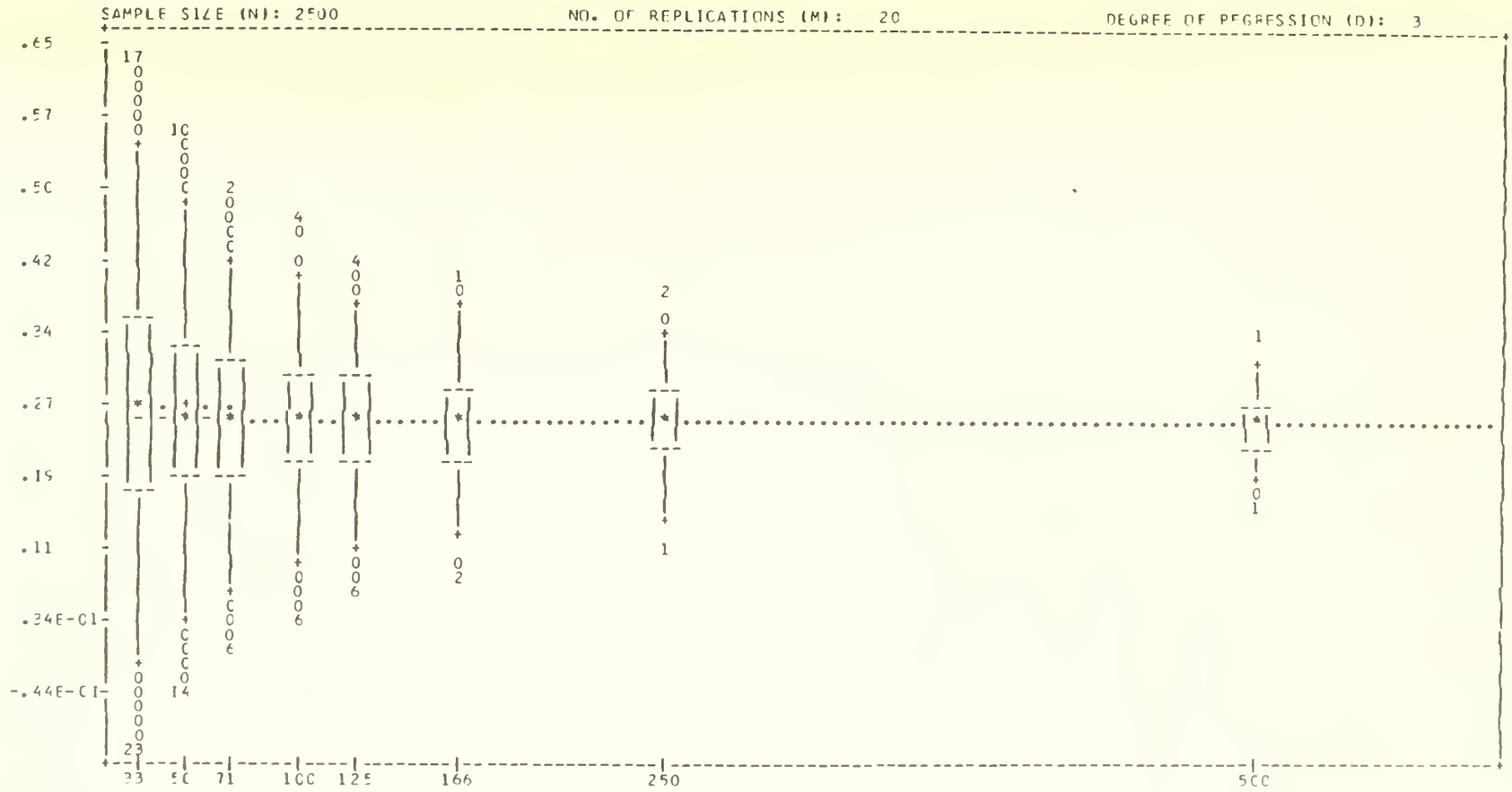
SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.2506	.2515	.2521	.2518	.2520	.2523	.2523	.2524
STD	.6316E-01	.4153E-01	.3564E-01	.2809E-01	.2575E-01	.2250E-01	.1818E-01	.1264E-01
STD MEAN	.1373E-02	.1313E-02	.1347E-02	.1256E-02	.1287E-02	.1322E-02	.1285E-02	.1264E-02
SKEWNESS	0.996	0.736	0.736	0.413	0.439	0.327	0.170	0.049
KURTOSIS	1.802	0.982	0.712	0.069	0.220	0.163	-0.322	-0.045

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.252594	-.924509E-01	4.61397	-122.667
VARIANCE OF REGRESSION - COEFFICIENTS:	.208554E-05	.181097E-01	116.478	54580.3
STD DEV OF REGRESSION - COEFFICIENTS:	.144414E-02	.134572	10.7925	233.624
REGRESSION ON VARIANCE - COEFFICIENTS:	.738654E-01	.237921	-2.13279	8.05335

VERTICAL SCALE: YMIN = -0.1055
YMAX = 0.6520

ESTIMATOR: 4-FOLD JACKNIFED MAXIMUM LIKELIHOOD ESTIMATE OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=.25

Figure 5(c)



SUBSAMPLE SIZE	33	50	71	100	125	166	250	500
MEAN	.2685	.2623	.2579	.2561	.2546	.2549	.2555	.2541
STD	.1594	.1225	.9679E-01	.8482E-01	.7370E-01	.6461E-01	.5092E-01	.3773E-01
STD MEAN	.4115E-02	.3081E-02	.2658E-02	.3793E-02	.3685E-02	.3731E-02	.3601E-02	.3773E-02
SKEWNESS	-0.158	-0.154	-0.204	-0.136	-0.103	-0.153	0.171	-0.014
KURTOSIS	0.952	0.550	0.451	0.684	0.435	0.401	0.301	1.329

MEAN OF REGRESSION ON AVERAGES - COEFFICIENTS:	.255242	-.374536	53.4121	-877.151
VARIANCE OF REGRESSION - COEFFICIENTS:	.146565E-04	.483583	3064.30	.152388E+07
STD DEV OF REGRESSION - COEFFICIENTS:	.382838E-02	.695401	55.3561	1234.46
REGRESSION ON VARIANCE - COEFFICIENTS:	.750446	-1.48800	4.12574	42.2654

VERTICAL SCALE: YMIN = -0.1055
YMAX = 0.6520

ESTIMATOR: 4-FOLD JACKNIFED MOMENT ESTIMATOR OF THE SHAPE PARAMETER OF THE GAMMA DISTRIBUTION K=.25

Figure 5(d)

DISTRIBUTION LIST

	NO. OF COPIES
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	2
Professor P. A. W. Lewis Code 55Lw Naval Postgraduate School Monterey, CA 93940	170